



# Recommender Workshop

## Part 2: Collaborative Filtering

# Recommender Workshop Agenda

- Part 1: Introduction
  - Overview of Machine Learning Process, Amazon SageMaker
  - Hands-on: **Data Exploration**
- **Part 2: Collaborative Filtering (You Are Here)**
  - Core Concepts for Recommendations
  - Hands-on: **K-Means Clustering**
- Part 3: Matrix Factorization
  - Refining Recommendations
  - Hands-on: **Factorization Machine**
- Part 4: Hyperparameter Optimization

# Recommendations

NEW & INTERESTING FINDS ON AMAZON [EXPLORE](#)

amazon prime  Prime Video Rent or Buy

Deliver to Steve  
Singapore 098657 Departments Browsing History Your Recommendations Today's Deals Gift Cards Registry Sell Help EN Hello, Steve Account & Lists Orders Prime Cart

1-24 of over 10,000 results for Electronics : Computers & Accessories : Computers & Tablets : Laptops : "laptop" Report a search issue

Show results for Any Category Electronics Computers & Accessories Computers & Tablets Laptops Traditional Laptops 2 in 1 Laptops

Refine by International Shipping  Ship to Singapore Amazon Prime  prime Laptop Display Size  17 Inches & Above  16 to 16.9 Inches  15 to 15.9 Inches  14 to 14.9 Inches  13 to 13.9 Inches  12 to 12.9 Inches  11 to 11.9 Inches  11 Inches & Under Computer RAM Capacity  64 GB & Above  32 GB  24 GB  16 GB

Sort by Relevance

 SPONSORED BY GOOGLE PIXELBOOK Save today on a Pixelbook Chromebook. Shop now

 Google Pixelbook (i5, 8 GB RAM, 128GB)  Google Pixelbook Pen  Google Pixelbook 45W USB Type-C Charger

 Sponsored by Samsung Chromebook 3, 11.6", 4GB RAM, 16GB eMMC, Chromebook (XE500C13-K04US)  \$200<sup>98</sup> \$219.99  prime Get it by Friday, May 11

 Sponsored by HP 2018 Flagship HP 15.6 Inch Touchscreen Notebook Laptop Computer (Intel Core i5-7200U 2.5GHz, 16GB DDR4 RAM, 512GB SSD, DVD, WiFi, Intel HD Graphics 620, HD Webcam, Windows 10) Silver  \$770<sup>00</sup>

Showing results in Electronics. Show instead results in All Departments.

Laptop Display Size: 17 Inches & Above | 16 to 16.9 Inches | 15 to 15.9 Inches | 14 to 14.9 Inches | 13 to 13.9 Inches | See more

Ad feedback

# Recommendations

prime

All laptop

Deliver to Steve  
Singapore 098657

Departments ▾ Browsing History ▾ Your Recommendations Today's Deals

1-24 of over 10,000 results for Electronics : Computers & Accessories : Computers & Tablets : Laptops : "laptop"

so many results

Show results for

- < Any Category
- < Electronics
- < Computers & Accessories
- < Computers & Tablets



SPONSORED BY GOOGLE PIXELBOOK

Save today on a Pixelbook Chromebook.

# SageMaker recommendation engine!

This is our demo of using SageMaker as recommendation engine for movies....

[Learn more »](#)

## Select movies you liked

 Type movies name to search

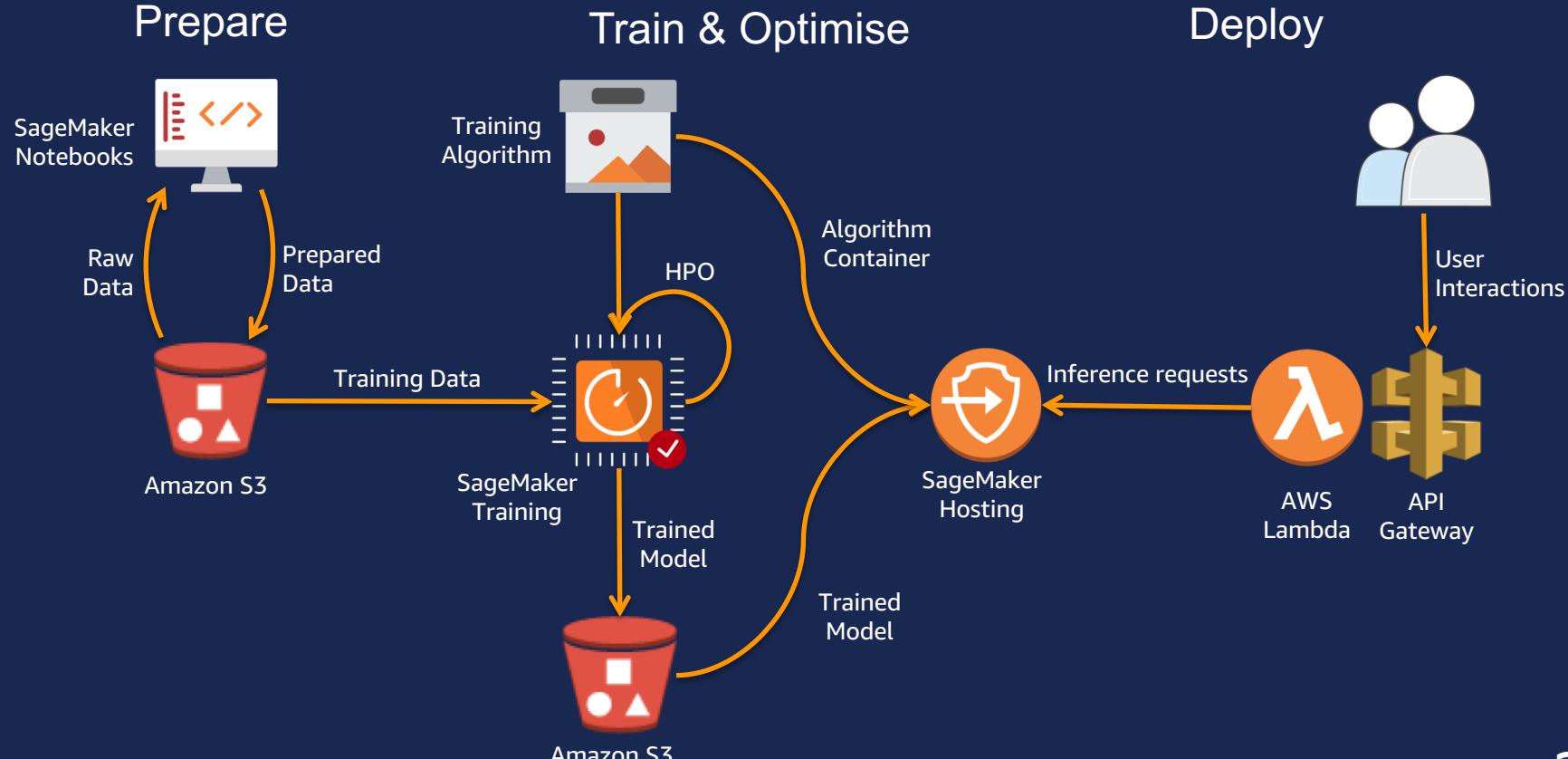
## Select recommendation engine

 Alastair's Beanstalk[View recommendation](#)

## Our recommendation



# Solution Architecture



# Recommenders: Use Case

## Product Matching

- Books, movies, music, etc based on personal preferences

## Online Ad Placement

- Identifying relevant products based on personal needs

# Recommender: Features

- What do we know about that user?
  - Demographics
  - Preferences/Behaviour
  - Networks
- What do we know about the products?
  - Category
  - Description/Text
  - Pricing/Sales History

# Recommender: Approach

- Find similar users
  - what are \*those\* individuals buying?
- Find similar products
  - what is associated with preferred products?

# Collaborative Filtering

# Collaborative Filtering: Defined

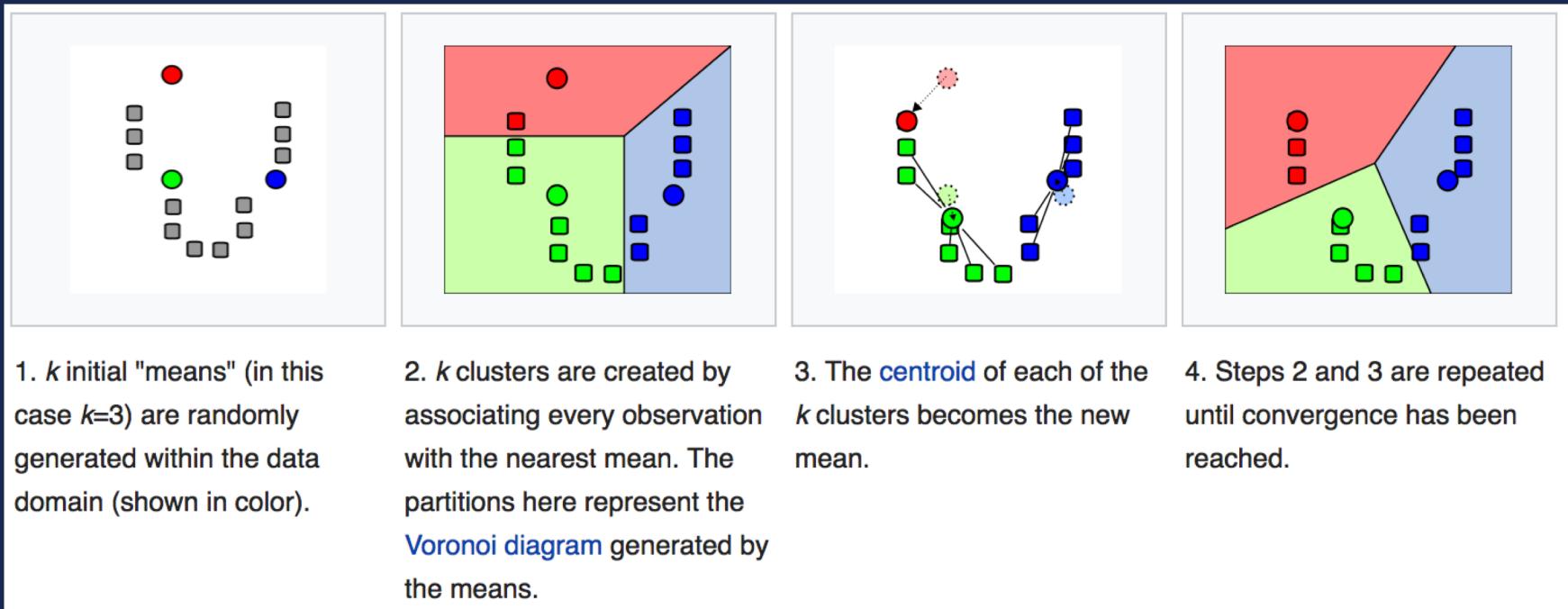
"[By mapping user preferences] users help one another to identify interesting items (although they are not aware of the collaboration), thus called collaborative filtering." [1]

[1] Agarwal, Deepak K.. Statistical Methods for Recommender Systems (p. 31). Cambridge University Press. Kindle Edition.

# Collaborative Filtering: Approach

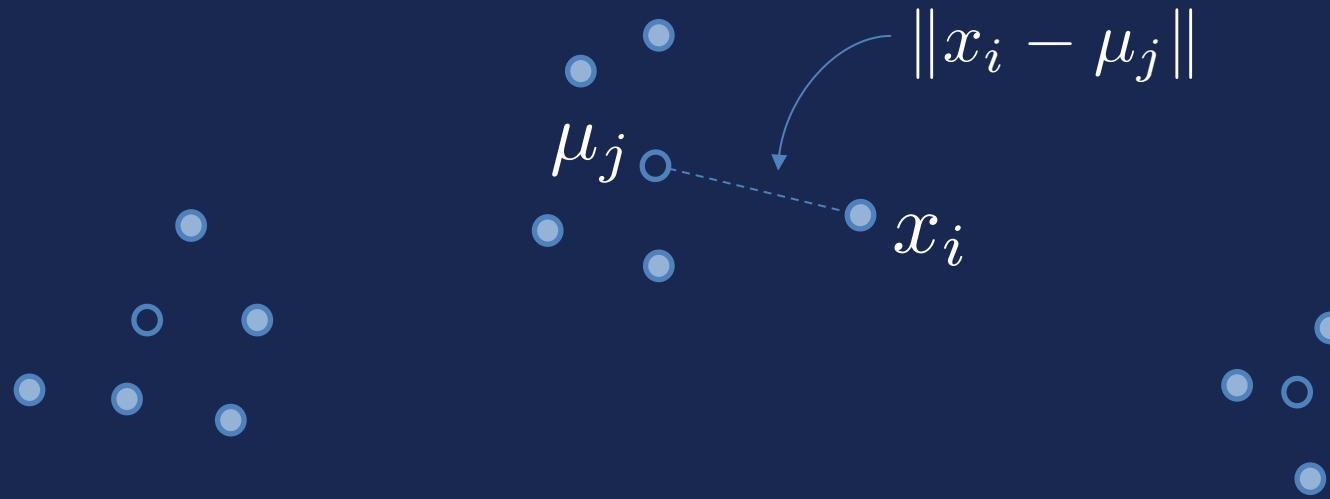
- Define similarity measures of interest
  - User-User
  - Item-Item
- Measure that similarity
  - K-Means Clustering

# K-Means Steps



[https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering)

# Amazon SageMaker: K-Means Clustering

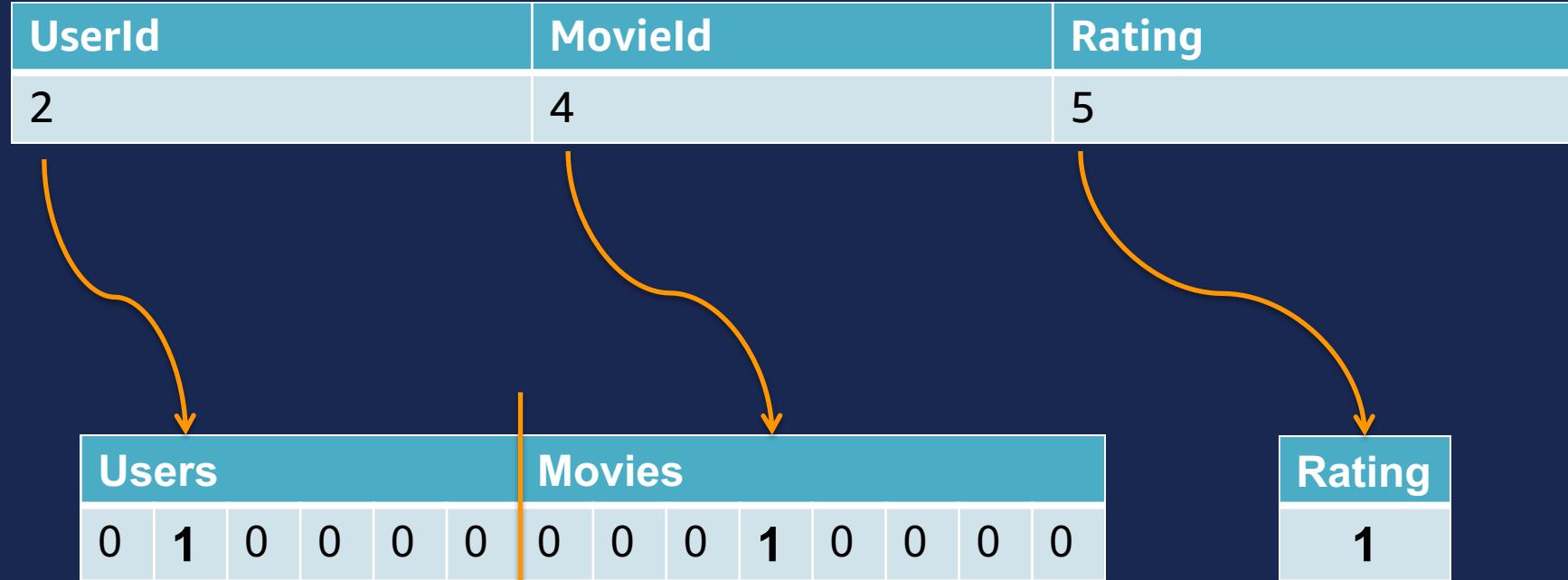


$$\frac{1}{n} \sum_i \min_j \|x_i - \mu_j\|^2$$

# Other Concepts: Hot and Cold

- Cold Start
- One-Hot Encoding

# Data Preparation: One-Hot Encoding



# Data Preparation: One-Hot Encoding

UserId	Movield	Rating
2	4	5
2	7	2

Users						Movies									Rating		
0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	
0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	

# Data Preparation: One-Hot Encoding

UserId	Movield	Rating
2	4	5
2	7	2
4	5	4

Users						Movies									Rating			
0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0

# Sparse Data

Users						Movies						Rating	
0	1	0	0	0	0	0	0	0	1	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	0	1	0	0	0

- One-Hot Encoding produces a a **2625x90,570** matrix for our training set
- This data set is **99.92% zeros**.
- Use a **memory efficient** data structure: **scipy.lil\_matrix**

# Recommender Workshop Repository

<http://bit.ly/2wkaV0N>

# Recommender Workshop Activity

- Log into the AWS console
- Change to **us-east-1** region
- Find the Amazon SageMaker service
- Find **Notebooks**
- Find your existing notebook instance (or create another instance)
  - Recommended: m1.m4 type
- Open the notebook instance and find within the repo path:
  - **02\_clustering\_users.ipynb**

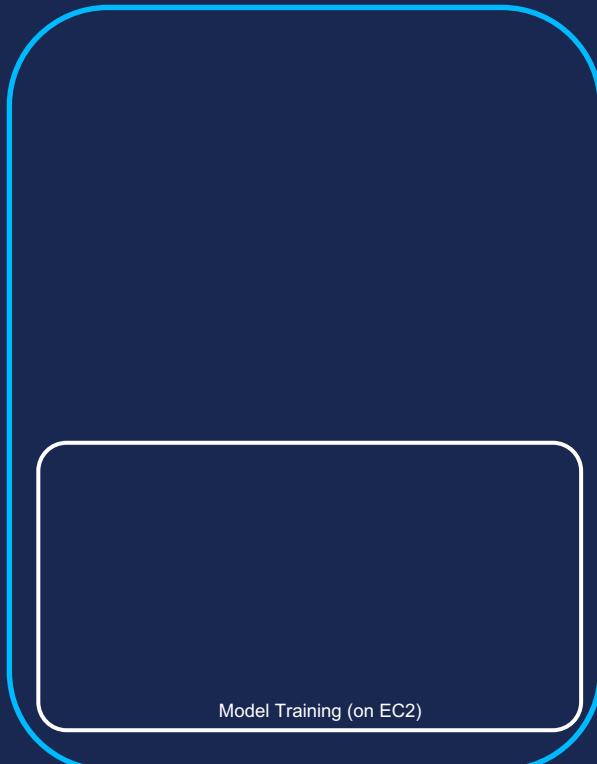
Client application

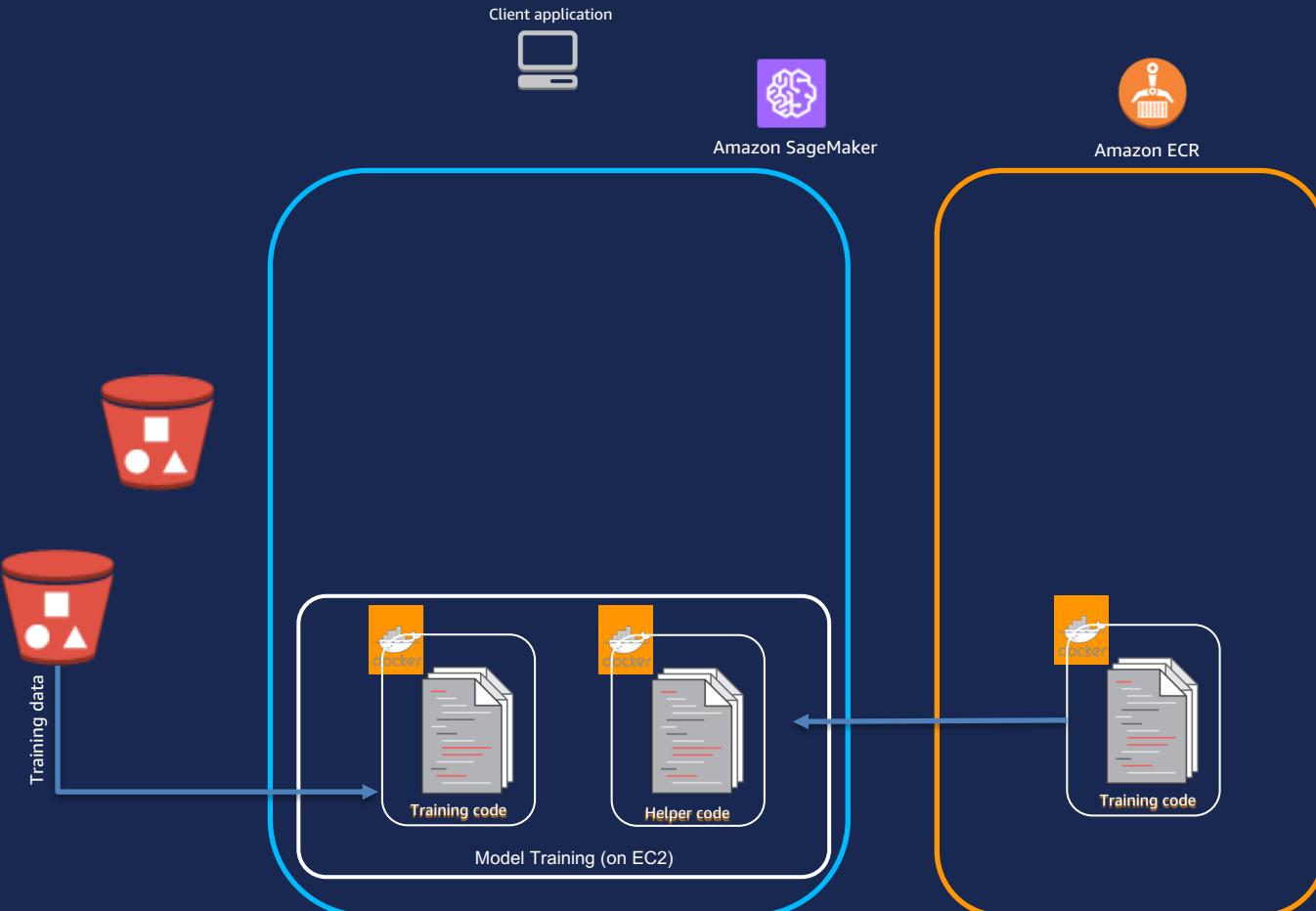


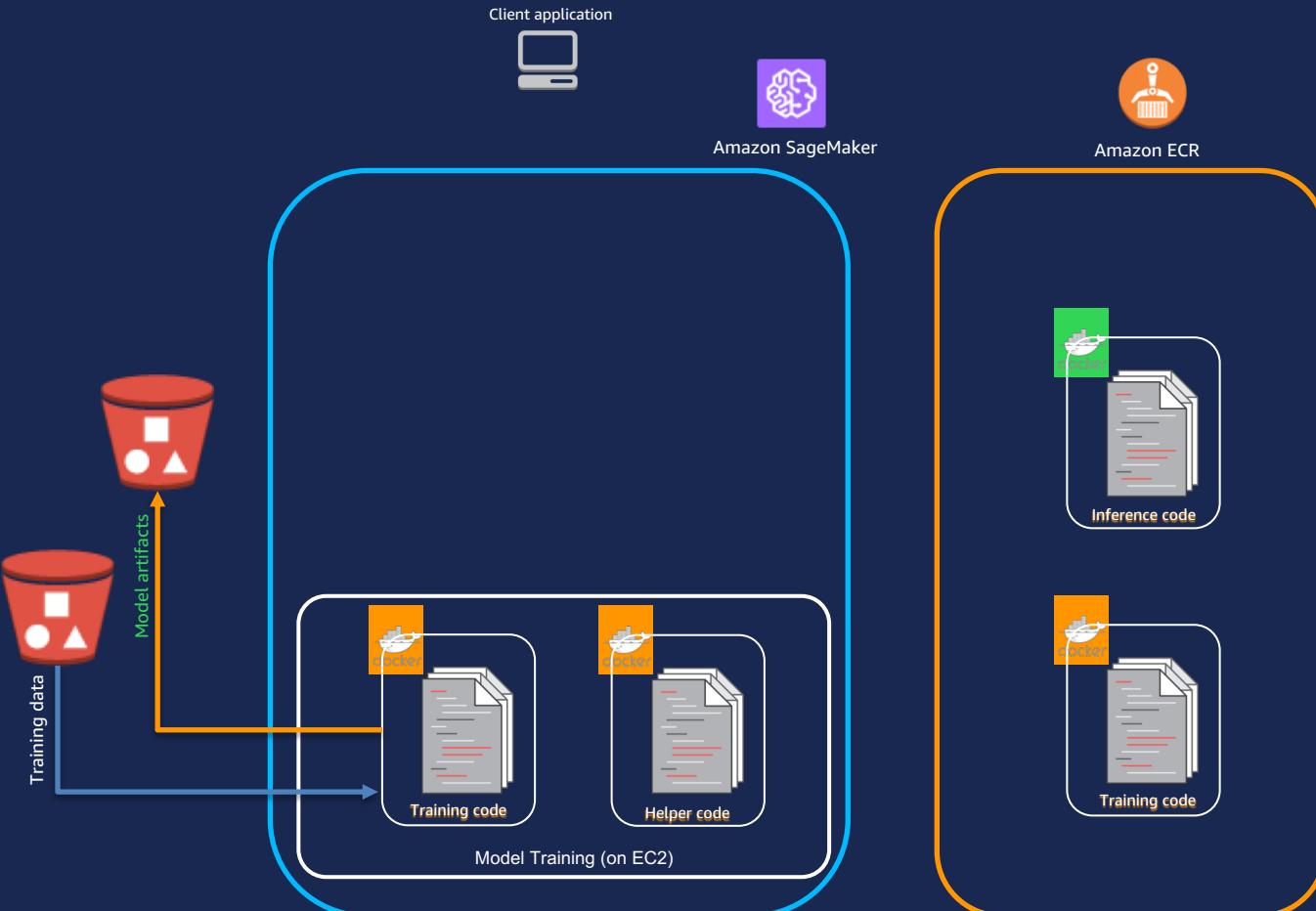
Amazon SageMaker

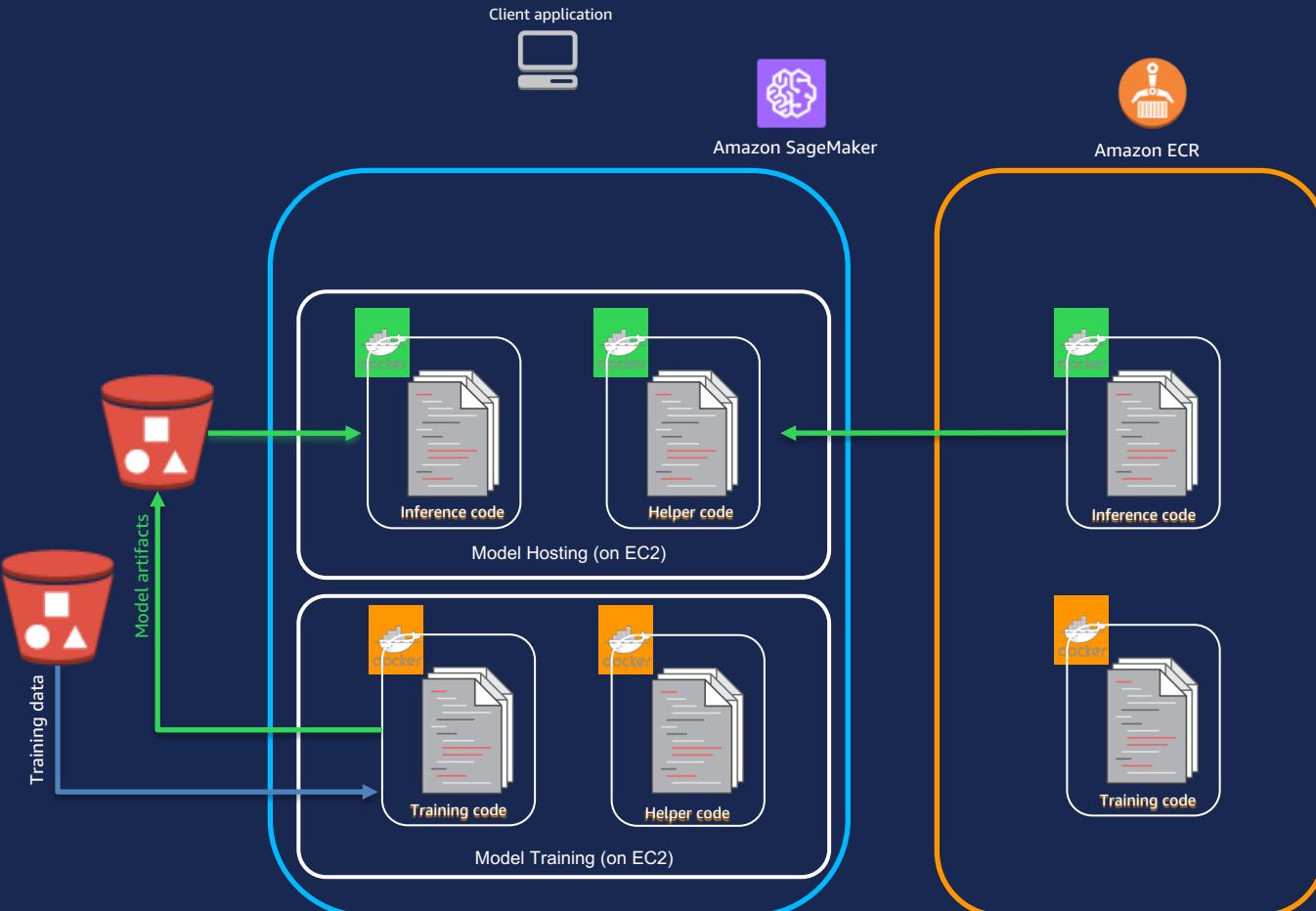


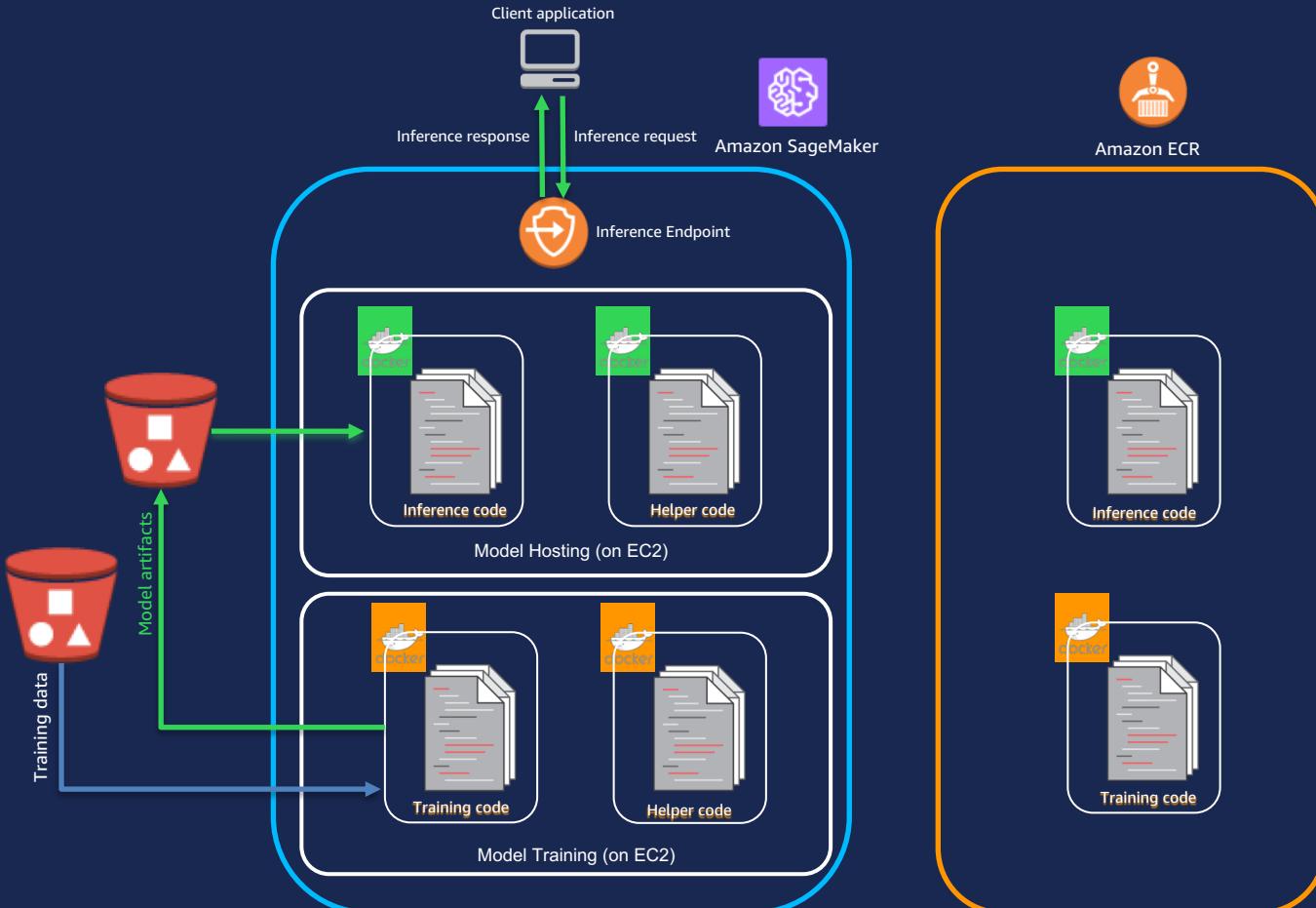
Amazon ECR

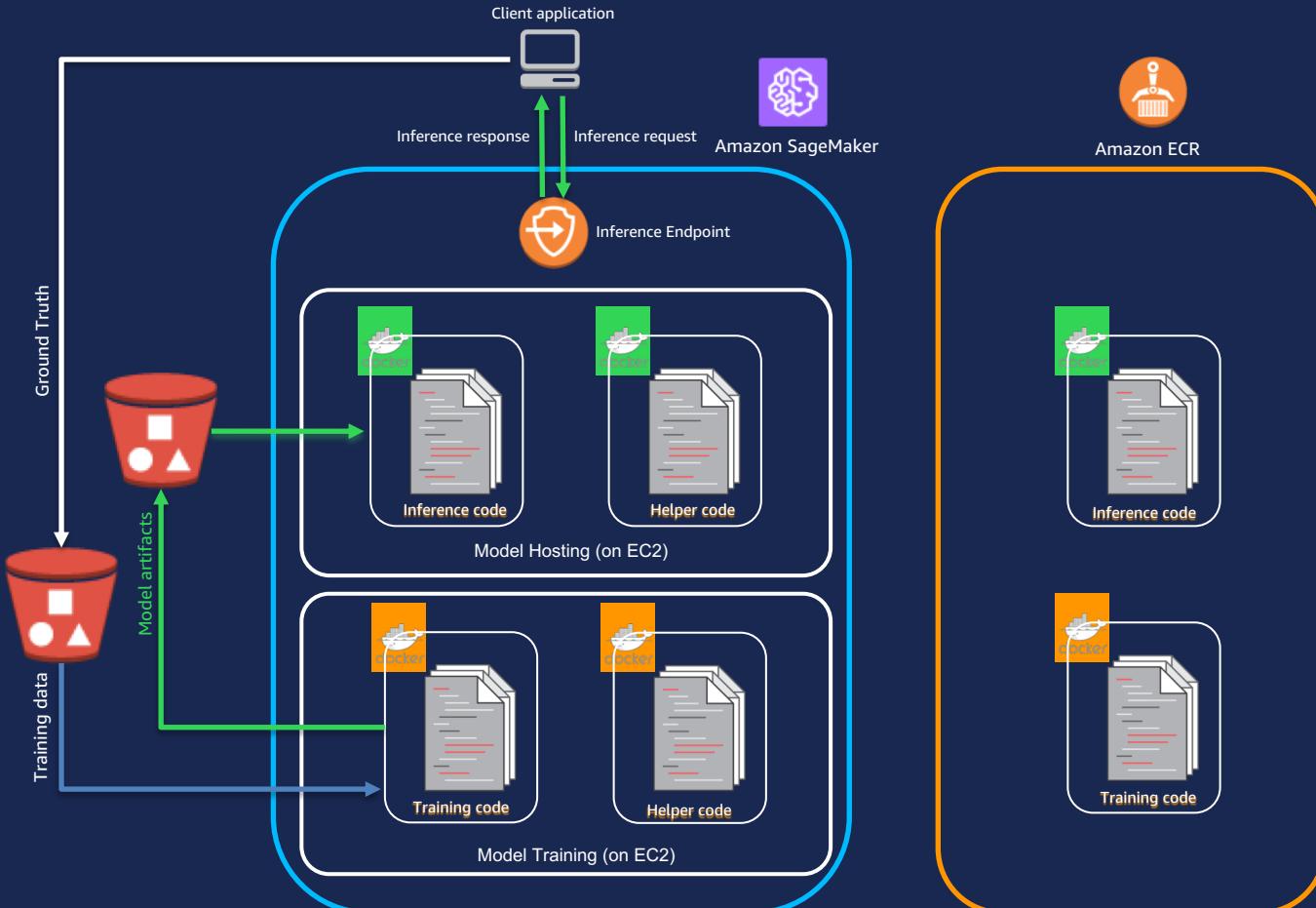












# Next: Part 3