



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 Tuning
 - Key Concepts
 - Hands-on: **Hyperparameter Tuning**

Recommendations

NEW & INTERESTING FINDS ON AMAZON **EXPLORE**

amazon prime All laptop

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

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


Google Pixelbook (i5, 8 GB RAM, 128GB)
4.5 stars 93 reviews
 prime


Google Pixelbook Pen
4.5 stars 39 reviews
 prime


Google Pixelbook 45W USB Type-C Charger
4.5 stars 1 review
 prime

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 ▾

Sponsored
Samsung Chromebook 3, 11.6", 4GB RAM, 16GB eMMC, Chromebook (XE500C13-K04US)
by Samsung
\$200⁹⁸ \$219.99 prime
Get it by Friday, May 11
4.5 stars 501 reviews



Sponsored
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
by HP
\$770⁰⁰
4.5 stars 7 reviews

Ad feedback

Recommendations

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

Show results for

so many results

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



SPONSORED BY GOOGLE PIXELBOOK

Save today on a Pixelbook
Chromebook.



Home TV Shows Movies Originals Recently Added My List

Search KIDS DVD Bell 4 Account

Trending Now



Oddballs & Outcasts ›



New Releases



Watch It Again



© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Customers who bought this item also bought

Page 1 of 5



Huggies Ultimate Sensitive Skin Baby Wipes Bundle Pack (Pack of 384)

★★★★★ 1

\$25.00



Huggies Fragrance Free Baby Wipes Bundle Pack (Pack of 480), 480 Wipes (6 x 80 Pack)

★★★★★ 1

\$25.00



Gumdrop Newborn Paci Boy 2pk 0m+

\$6.99



Huggies Fragrance Free Baby Wipes (Pack of 320), 320 Wipes (4 x 80 Pack)

★★★★★ 1

\$17.33



Huggies Aloe & Cucumber Baby Wipes (Pack of 320), 320 Wipes (4 x 80 Pack)

\$13.33



Curash Water Baby Wipes 80 PK

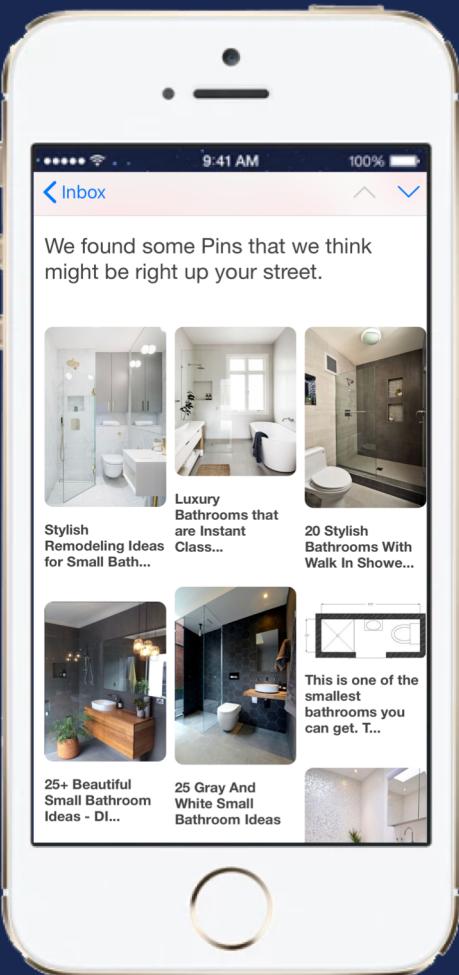
\$4.99



Huggies Essentials Nappies, Size 2, 54 Count

\$13.00

Pinterest



© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws
amazon

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.

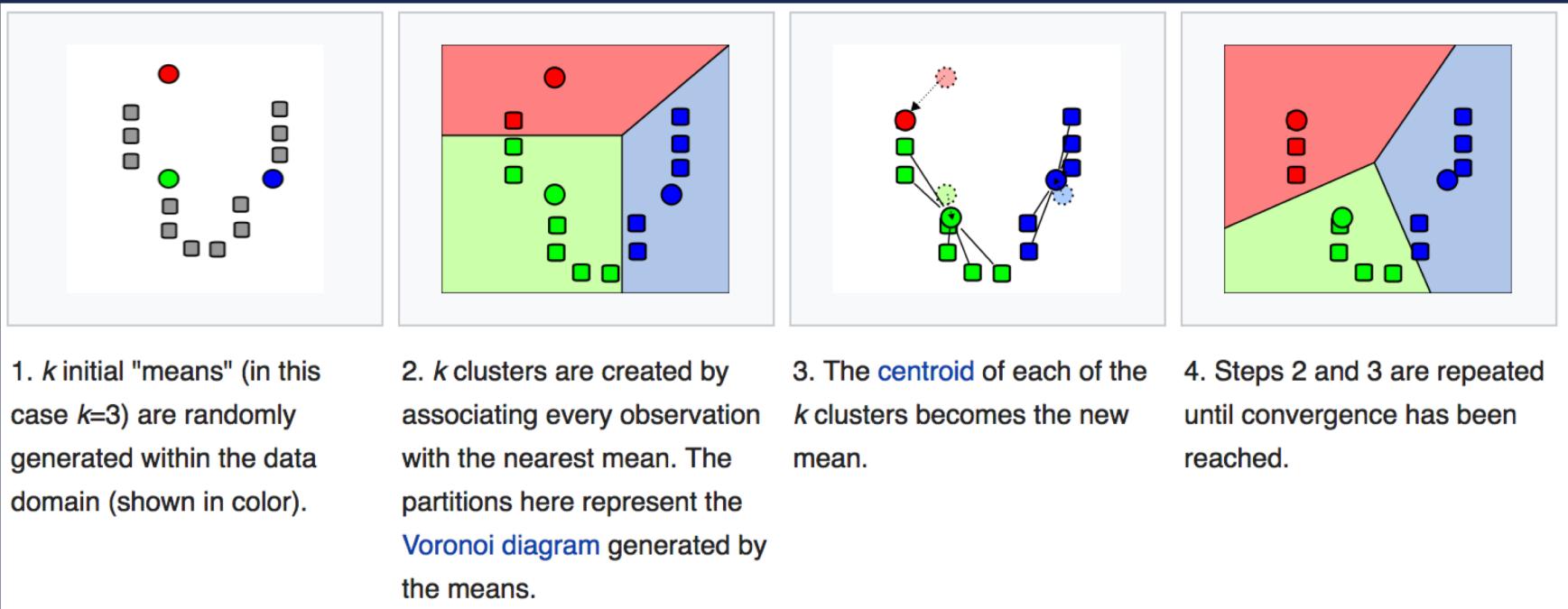
Collaborative Filtering: Why?

- Define similarity measures of interest
 - User-User
 - Item-Item

Why User-User?

- Scale
- Cold-Start problems
- Simplifying the problem space

How: K-Means



https://en.wikipedia.org/wiki/K-means_clustering

Data Preparation

Our Data Set: MovieLens

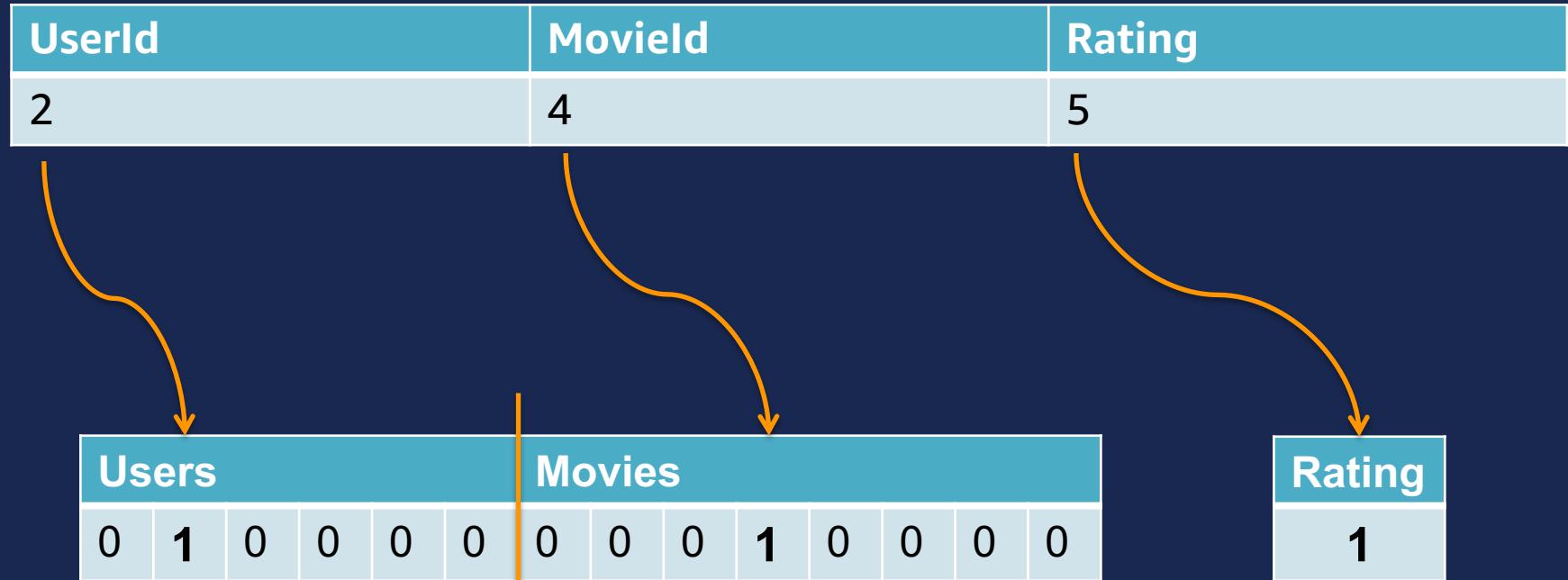
- Public Data Set produced by **GroupLens Research**
- <https://grouplens.org/datasets/movielens/>

```
In [15]: data = pd.read_csv("u.data", sep='\t', header=None,
                           names=['userid', 'movieid', 'rating', 'timestamp'])
data.head()
```

Out[15]:

	userid	movieid	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

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	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	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	0	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	0	0	0

Sparse Data

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

- 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 Activity

- Log into the AWS console
- Find the Amazon SageMaker service
- Find Notebooks
- Find your existing notebook instance (or create another instance)
 - Recommended: ml.m4 type
- Open the notebook instance and find within the repo path:
 - `02_clustering_users.ipynb`

Wait.... What did we just do?

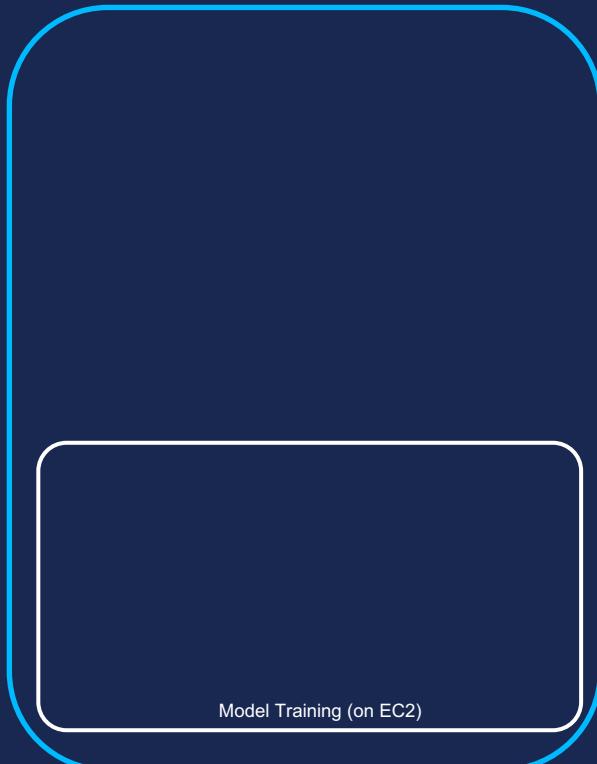
Client application

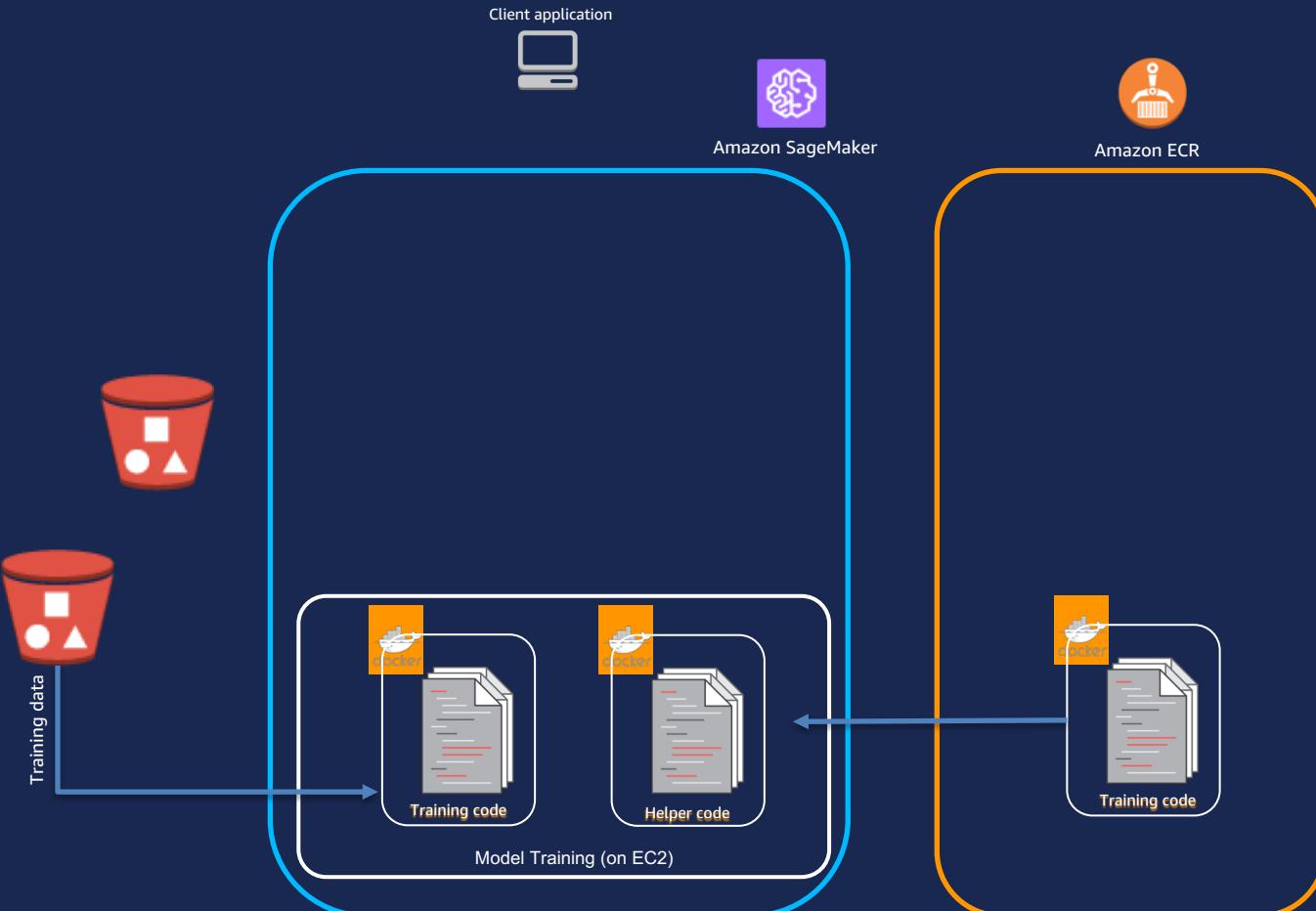


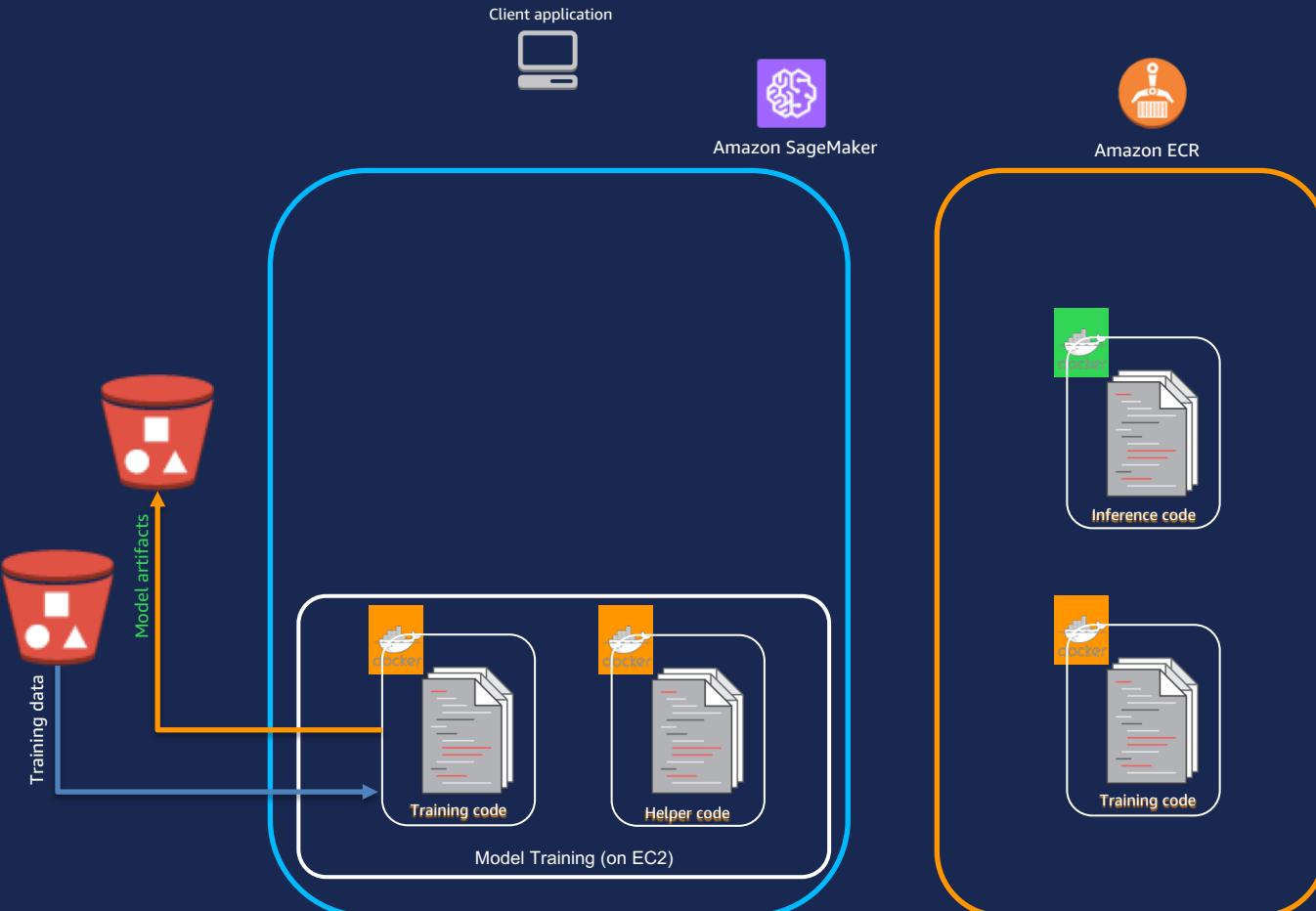
Amazon SageMaker

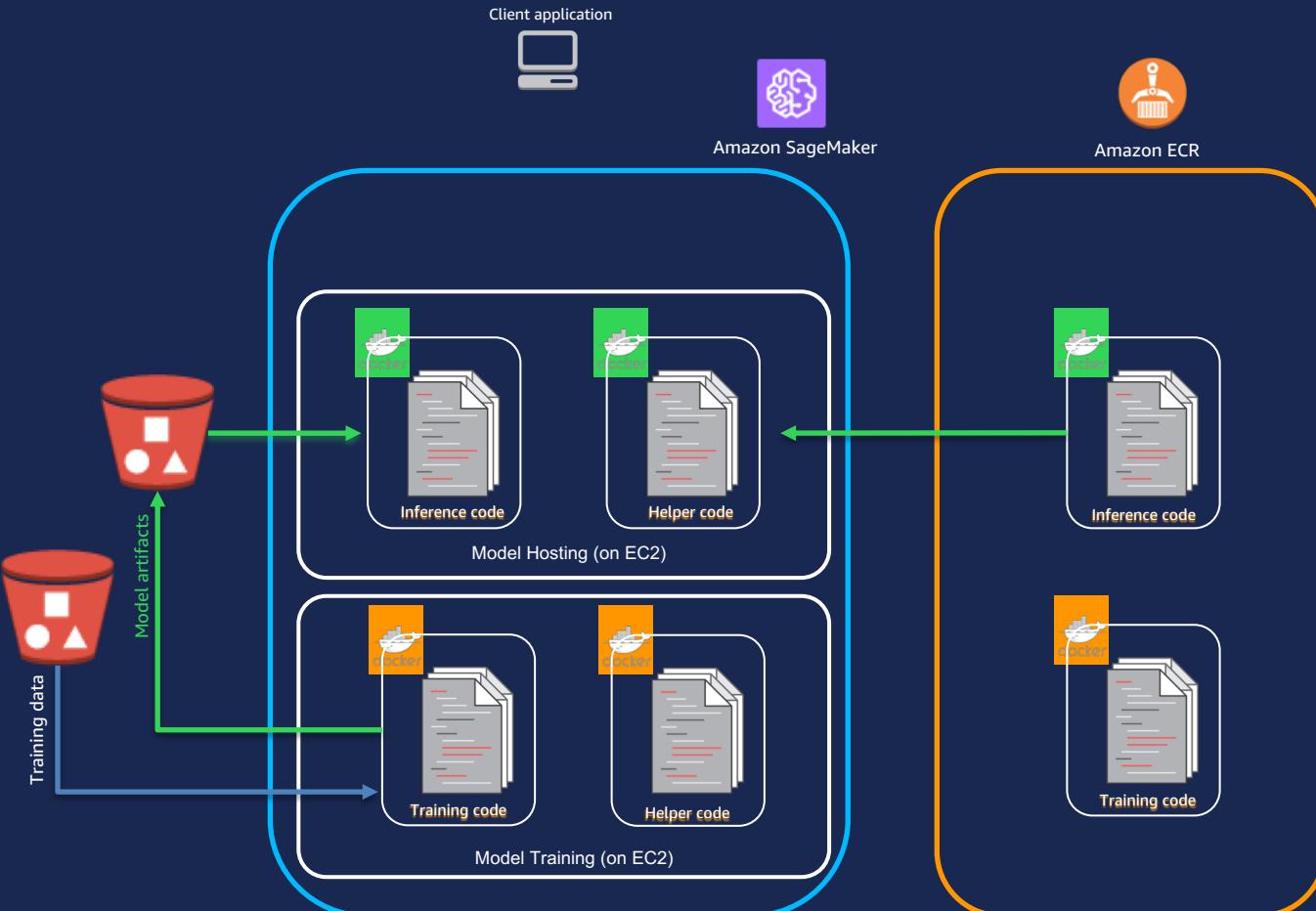


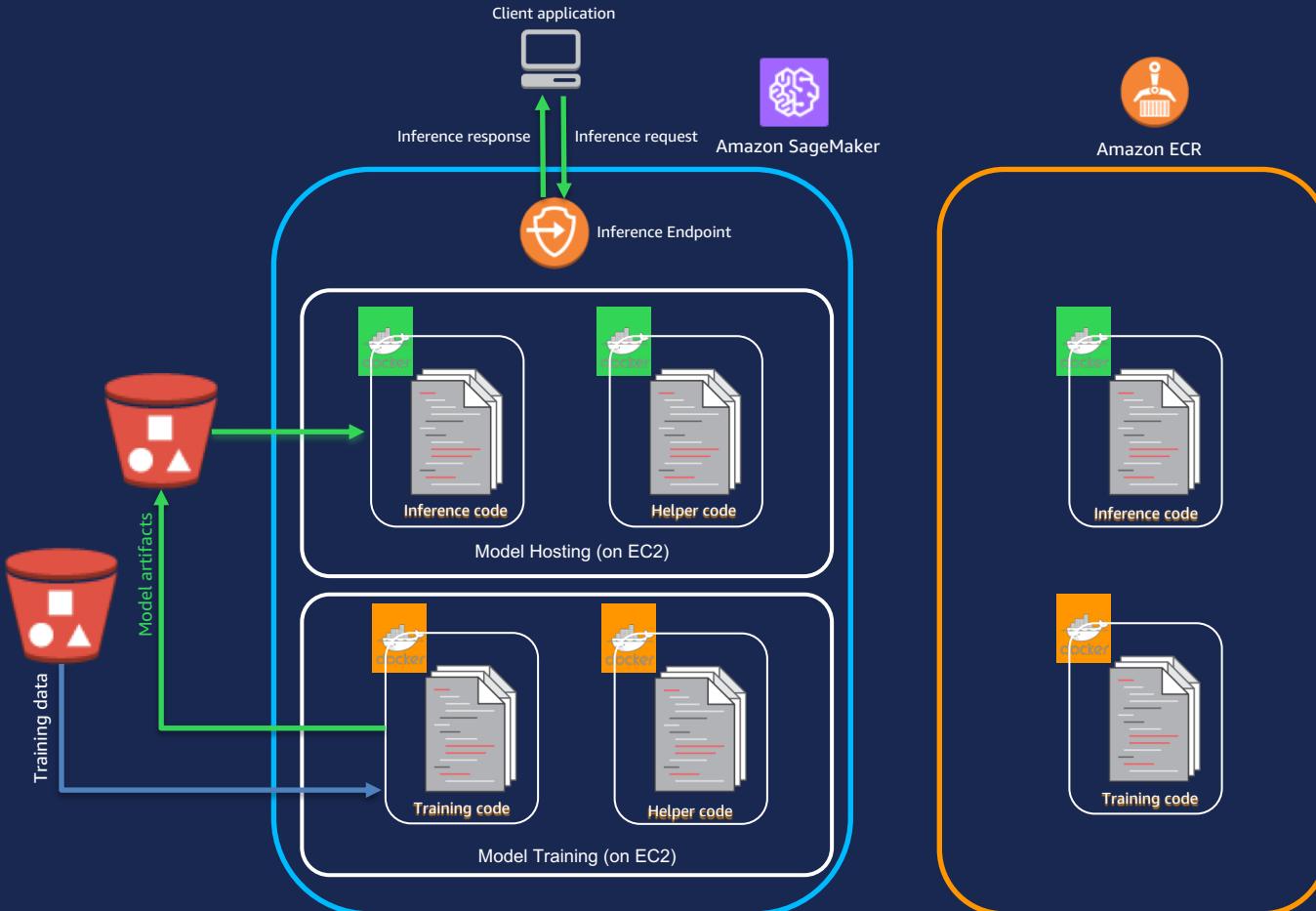
Amazon ECR

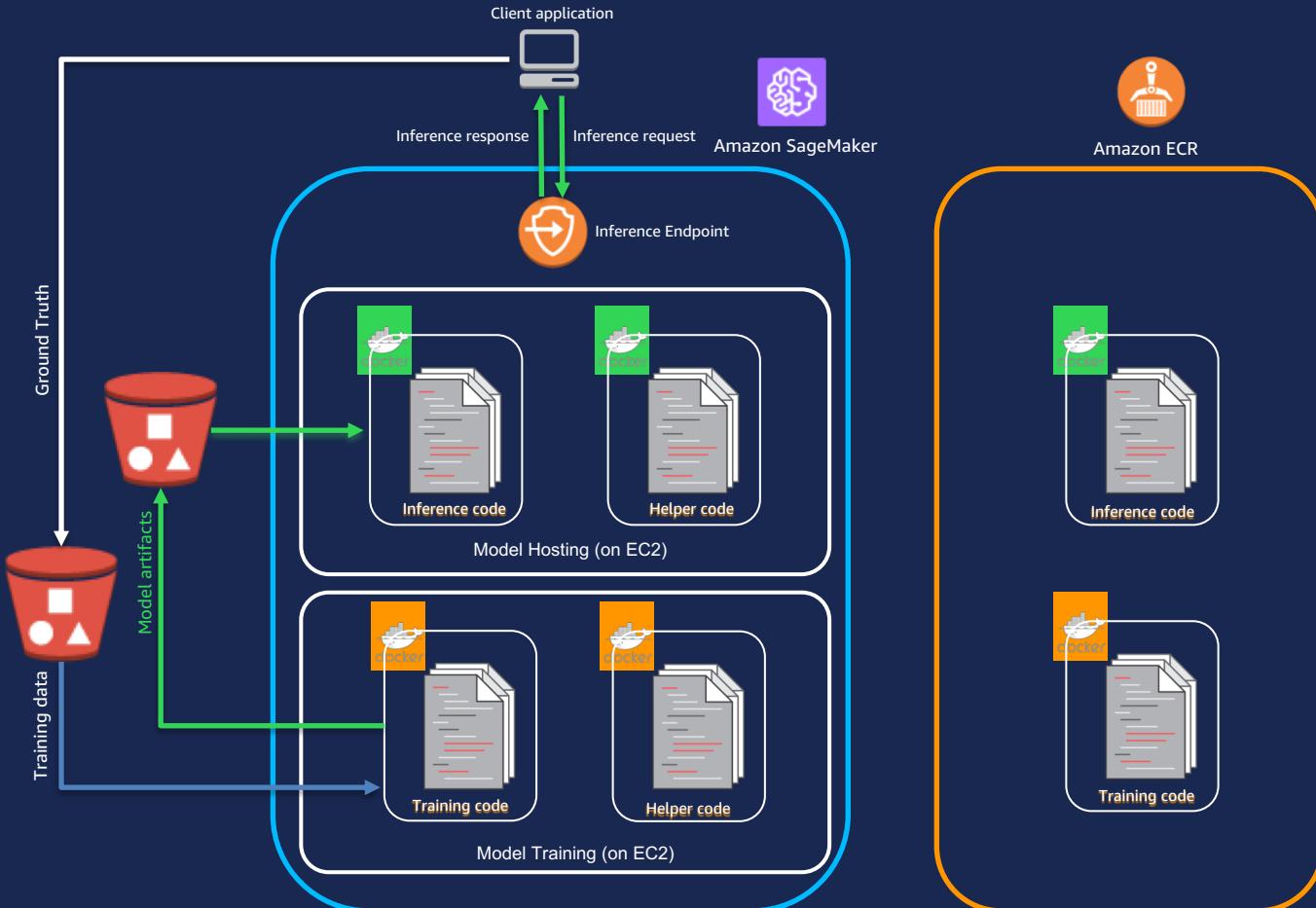












Next: Part 3