



SageMaker Low-Code Machine Learning

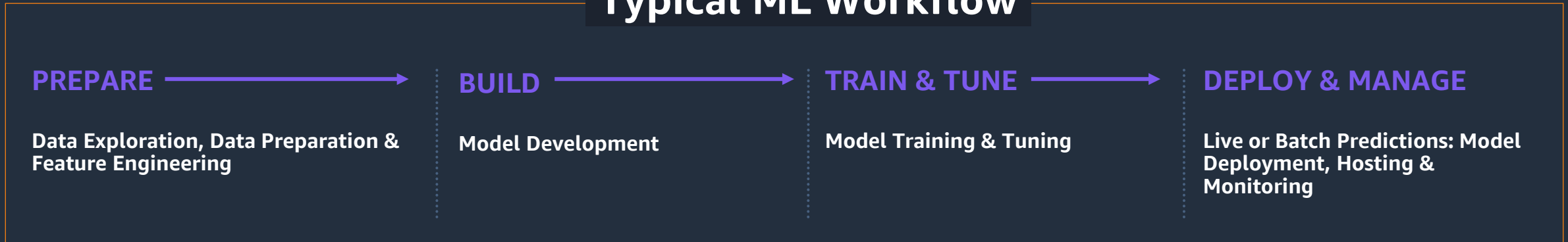
Accelerate your Machine Learning Workflows

Charles Laughlin
Principal AI/ML Solutions Architect, AWS

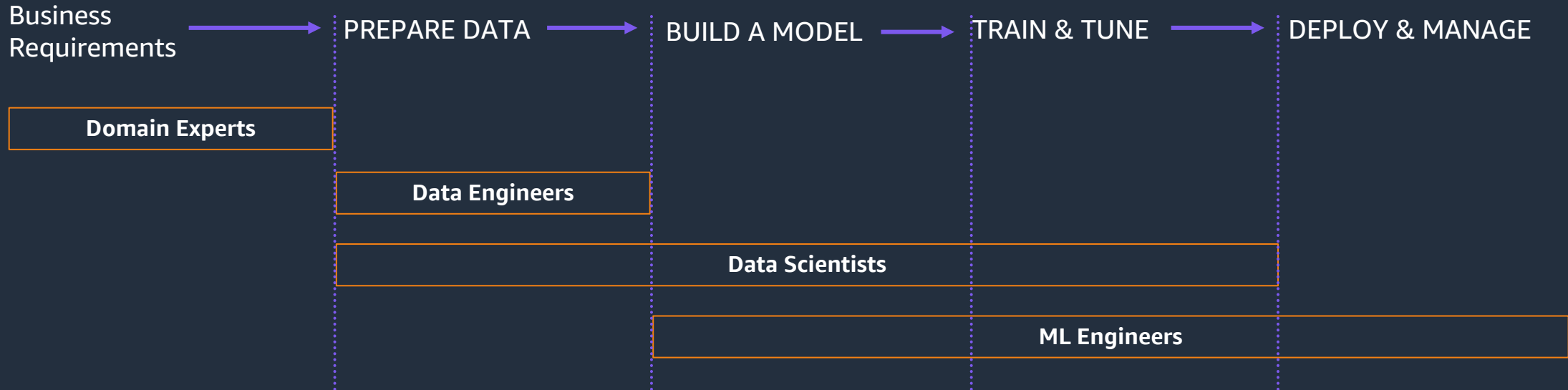
Claire O'Brien Rajkumar
Sr. Product Manager, AWS

The Machine Learning Workflow

Typical ML Workflow



ML WORKFLOW IN AN ORGANIZATION



Usually takes from weeks to months primarily solving for the prioritized use cases

Low-Code Machine Learning on AWS

Amazon SageMaker

PREPARE

Data Exploration, Data Preparation & Feature Engineering

Amazon SageMaker Data Wrangler

A faster, visual way to aggregate & prepare data for machine learning

BUILD

Model Development

Amazon SageMaker Autopilot

AutoML capability that automatically prepares your data, as well as builds, trains, & tunes the best machine learning models for your tabular datasets

Amazon SageMaker Canvas

A visual point-and-click interface to generate accurate ML predictions— without requiring any machine learning experience or having to write a single line of code.

Amazon SageMaker JumpStart

A model hub of pre-trained and easily tunable state-of-the-art models for Computer Vision, & Natural Language Processing tasks

TRAIN & TUNE

Model Training & Tuning

DEPLOY & MANAGE

Live or Batch Predictions: Model Deployment, Hosting & Monitoring

Low-Code Machine Learning



Faster ML workflows



Open and customizable solutions

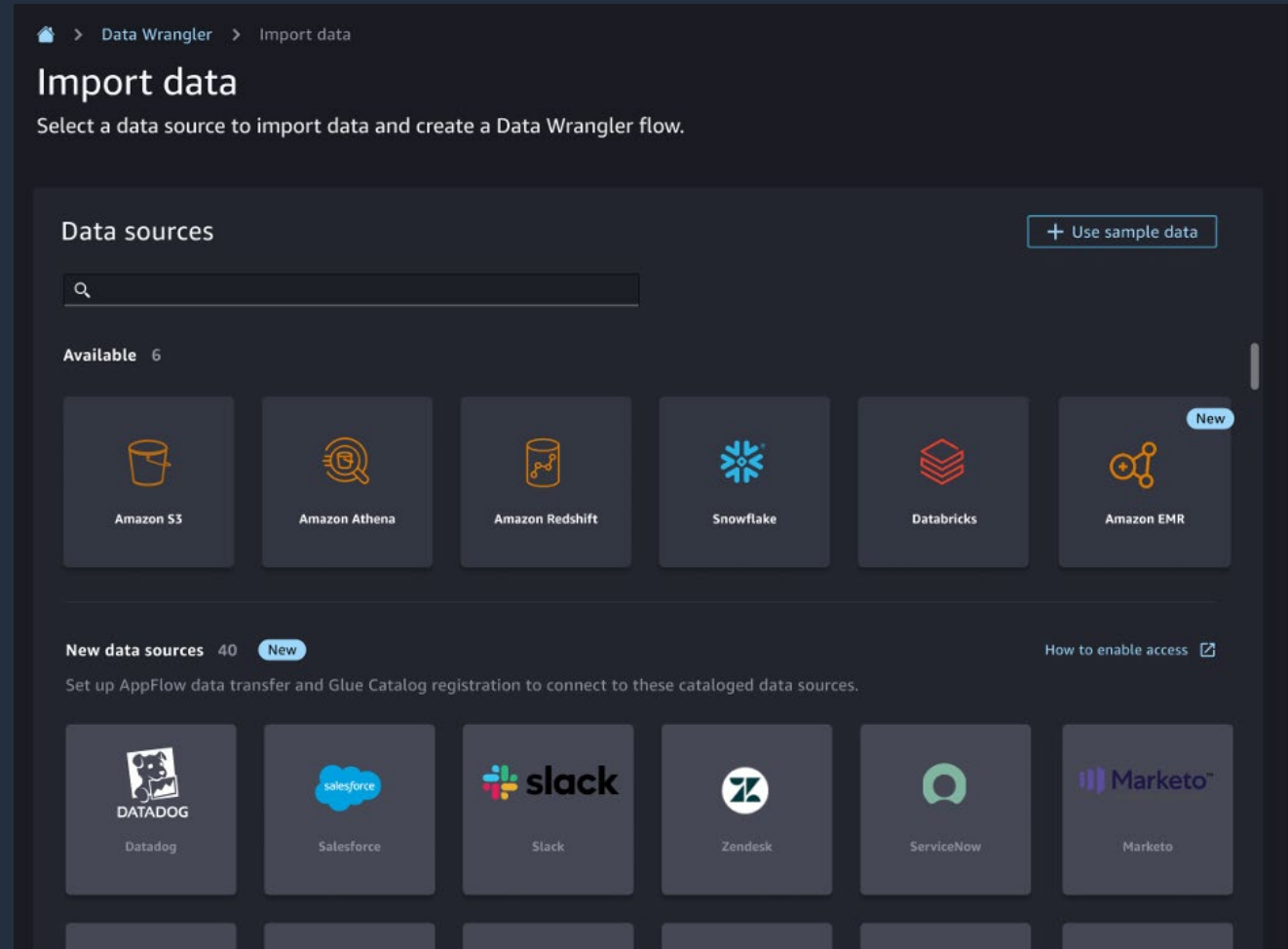


Appropriate for both new and advanced ML practitioners

SageMaker Data Wrangler – Product Updates

New Data Sources

- 40+ SaaS Data Sources via Appflow
- EMR
- Snowflake oAuth and direct data connection



SageMaker Data Wrangler – Product Updates

New Data Types

- Support for image data
- Support for text data (via custom transforms)

The screenshot displays the SageMaker Data Wrangler interface. On the left, a grid of image thumbnails is shown under the heading "Data" and "Step 1. S3 Source". Each thumbnail includes a filename (e.g., 1255.jpg, 13383.jpg, 13238.jpg, 10328.jpg, 10450.jpg, 11166.jpg, 112, 10920.jpg, 11777.jpg, 117) and a resolution label (e.g., H:311 W:415 C:3). On the right, a "CUSTOM TRANSFORM" dialog box is open, showing a Python (PySpark) script for text processing. The script includes imports for nltk and pyspark.sql, and defines a UDF to remove stopwords from text.

```
1  
2 from nltk import download  
3 from nltk.corpus import stopwords  
4 from nltk.tokenize import word_tokenize  
5 from pyspark.sql.functions import col, udf  
6  
7 download('stopwords')  
8 download('punkt')  
9 stop_words = set(stopwords.words('english'))  
10  
11 def remove_stop_words(str_in):  
12     word_tokens = word_tokenize(str_in)  
13     filtered_sentence = [w for w in word_tokens if not w.lower() in stop_words]  
14     return " ".join(filtered_sentence)  
15  
16 convertUDF = udf(lambda z: remove_stop_words(z))  
17  
18 df = df.select(col("label"), \n/>19               convertUDF(col("text")).alias("text") )
```

SageMaker Data Wrangler – Product Updates

Notebook widget

- Low-code data prep for notebooks
- 1-click to add code-based transforms

The screenshot displays the SageMaker Data Wrangler interface within a Jupyter Notebook. The top panel shows a code cell with the command `import pandas as pd`. Below the code, a widget titled "Clean data in minutes" offers to automatically visualize data and improve quality. The interface features several data visualization panels: a "ZIP CODE" panel with a bar chart and a "COMPANY" panel with a bar chart. A central table displays data for various hotel categories, including Hotel Room, Data Package, and Total Book. The bottom panel shows a code cell with `df = pd.read_csv('hotel_bookings_dirty_v2.3.csv')` and a warning message about mixed data types. Below the code, a "View Pandas table" button is visible, along with a table of data rows and a "DESCRIPTION" tooltip for the "company" column. The tooltip text reads: "Replaces missing values with the string that has the following value: 'Other' if textual data and 0 if numerical data". To the right, a "COMPANY" insights panel shows "Missing values: High" and "SUGGESTED TRANSFORMS" such as "Replace with new value" and "Drop column".

Check out our self-guided onboarding tutorials at
<https://aws.amazon.com/sagemaker/getting-started/>

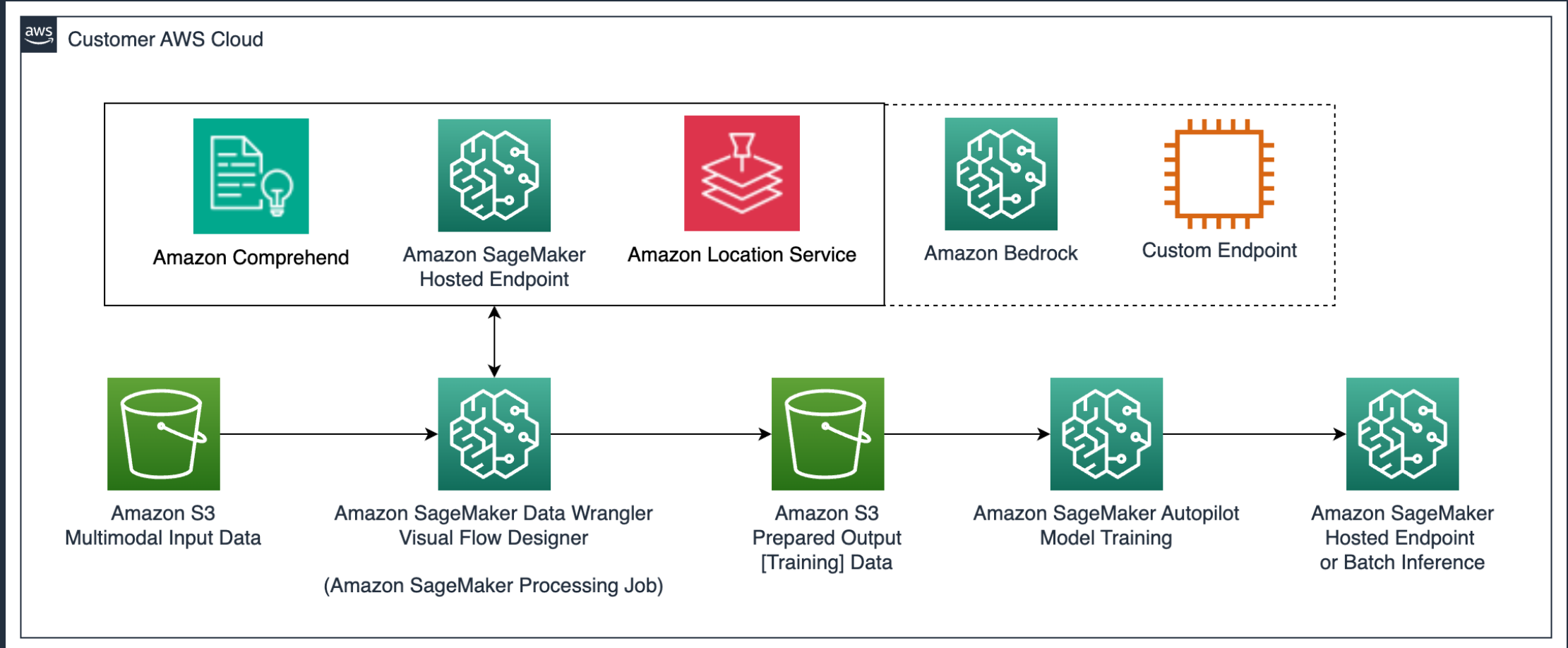


Using FM to modernize housing price estimation



Examples of kitchen settings; photos credit Francesca Tosolini (L) and Sidekix Media (R) on Unsplash

Incorporating diverse FM into data preparation workflows



Using FM to modernize housing price estimation



What are the countertops made of?
marble

Is this a large kitchen?
yes

Is this an expensive kitchen?
yes

How wide is the refrigerator?
4 feet



Examples of kitchen settings; photos credit Sidekix Media (R) on Unsplash

SageMaker AutoPilot– Product Updates

New Training Modes

- Support for Ensemble training mode
- Up to 8x Faster

Create an Autopilot experiment

When you create an Autopilot experiment, Amazon SageMaker analyzes your data and creates a notebook with candidate model definitions. This notebook provides visibility into how models are selected, trained, and tuned.

Experiment and data details

Training method

Deployment and advanced settings

Review and create

Training method

Select the training method for solving your machine learning problems.

Auto

Let Autopilot automatically decide the training method based on your dataset size.

Ensembling

Autopilot uses an AutoML algorithm that trains a multi-layer stack ensemble model to predict on regression and classification datasets directly from your data.

Hyperparameter optimization (HPO)

Autopilot finds the best version of a model by tuning hyperparameters and running training jobs on your data set.

Cancel

Previous: Experiment and data details

Next: Deployment and advanced settings

Low-Code Machine Learning on AWS

Amazon SageMaker

PREPARE

Data Exploration, Data Preparation & Feature Engineering

Amazon SageMaker Data Wrangler

A faster, visual way to aggregate & prepare data for machine learning

BUILD

Model Development

Amazon SageMaker Autopilot

AutoML capability that automatically prepares your data, as well as builds, trains, & tunes the best machine learning models for your tabular datasets

Amazon SageMaker Canvas

A visual point-and-click interface to generate accurate ML predictions— without requiring any machine learning experience or having to write a single line of code.

Amazon SageMaker JumpStart

A model hub of pre-trained and easily tunable state-of-the-art models for Computer Vision, & Natural Language Processing tasks

TRAIN & TUNE

Model Training & Tuning

DEPLOY & MANAGE

Live or Batch Predictions: Model Deployment, Hosting & Monitoring


SageMaker Canvas– Product Updates

New Use Cases

- NLP
- CV

Problem type

Select the problem type you want the model to solve.



Predictive analysis

Build models using tabular datasets to predict single or multiple categories as well as regression and time-series forecast problems.





Image analysis

Build models using image datasets to predict single or multiple categories for image classification problems.



Text analysis

Build models using tabular datasets to predict single or multiple categories for text classification problems.

[Cancel](#) [Create](#)

SageMaker Canvas– Product Updates

New Integrations

- Visualize predictions in Amazon Quicksight
- Share model to SageMaker Studio and deploy to end point

