# Steps to Use Pre-built Docker Image for Custom Training Job in Google Cloud Vertex AI

## High-Level Steps

1. Prepare Your Training Script: Write your training script (e.g., train.py) and upload it to a Google Cloud Storage bucket.

Note: Training Script is the components of the notebook where you run your pre-processing and define what your model will do.

2. Set Up YAML Configuration: Create a YAML configuration file to define your training job.

3. Submit the Training Job: Use the gcloud CLI to submit the training job with the specified configuration.

## Detailed Steps and Commands

### 1. Prepare Your Training Script

Write your training script and save it locally as train.py. Upload the script to a Cloud Storage bucket using the following command:

*gsutil cp train.py gs://your-bucket-name/train.py*

### 2. Set Up YAML Configuration

Create a file named config.yaml with the following content:

*trainingInput:  
 scaleTier: CUSTOM  
 masterType: n1-standard-4  
 masterConfig:  
 imageUri:* ***us-docker.pkg.dev/vertex-ai/training/tf-cpu.2-14.py310:latest*** *args:  
 - --data-path=gs://your-bucket-name/path-to-your-data.csv  
 - --bucket-name=your-bucket-name  
 - --model-path=model-outputs/model.pkl*

Replace your-bucket-name, path-to-your-data.csv, and model-outputs/model.joblib with your specific paths and filenames.

### 3. Submit the Training Job

Use the following gcloud command to submit your training job to Vertex AI:

*gcloud ai custom-jobs create \*

*--region=us-central1 \*

*--display-name=your-job-name \*

*--config=config.yaml*

## Summary of Bash Commands

Upload training script to Cloud Storage:

*gsutil cp train.py gs://your-bucket-name/train.py*

Submit the training job to Vertex AI:

*gcloud ai custom-jobs create \*

*--region=us-central1 \*

*--display-name=your-job-name \*

*--config=config.yaml*

## Summary of YAML Configuration

*trainingInput:  
 scaleTier: CUSTOM  
 masterType: n1-standard-4  
 masterConfig:  
 imageUri: us-docker.pkg.dev/vertex-ai/training/tf-cpu.2-14.py310:latest  
 args:  
 - --data-path=gs://your-bucket-name/path-to-your-data.csv  
 - --bucket-name=your-bucket-name  
 - --model-path=model-outputs/model.joblib*

Further Documentation:

[Available Pre-built Images that you can leverage](https://cloud.google.com/vertex-ai/docs/training/pre-built-containers#tensorflow)

[Creating your image](https://cloud.google.com/vertex-ai/docs/training/create-custom-container)

[Pushing and Submitting a Custom Training Job](https://cloud.google.com/vertex-ai/docs/training/create-custom-job)

[End-to-End Tutorial](https://cloud.google.com/vertex-ai/docs/pipelines/continuous-training-tutorial) (It's technically dense so feel free to take your time as you get through it).

Questions: Fedaarroyo@gmail.com