**pipeline.yaml**

apiVersion: tekton.dev/v1

kind: Pipeline

metadata:

name: simple-ci-pipeline

spec:

workspaces:

- name: shared-workspace

tasks:

# Step 1: Install dependencies

- name: install-requirements

taskSpec:

workspaces:

- name: src

steps:

- name: pip-install

image: python:3.10

env:

- name: HOME

value: /tekton/home

workingDir: $(workspaces.src.path)

script: |

set -euo pipefail

python --version

pip install --upgrade pip

pip install -r requirements.txt

workspaces:

- name: src

workspaces:

- name: src

workspace: shared-workspace

# Step 2: Run model training with spark-submit

- name: run-training

runAfter: [install-requirements]

taskSpec:

workspaces:

- name: src

steps:

- name: spark-train

image: gcr.io/spark-operator/spark-py:v3.1.1 # PySpark-enabled container

workingDir: $(workspaces.src.path)

script: |

set -euo pipefail

echo "Launching Spark training job..."

spark-submit \

--py-files ml\_stuff.zip \

model\_training.py

workspaces:

- name: src

workspaces:

- name: src

workspace: shared-workspace

**pipelinerun.yaml**

apiVersion: tekton.dev/v1

kind: PipelineRun

metadata:

generateName: ci-sandbox-run-

spec:

pipelineRef:

name: simple-ci-pipeline

timeouts:

pipeline: 30m # give more time for Spark training

workspaces:

- name: shared-workspace

volumeClaimTemplate:

spec:

accessModes: ["ReadWriteOnce"]

resources:

requests:

storage: 2Gi