**Pipeline.yaml**

apiVersion: tekton.dev/v1

kind: Pipeline

metadata:

name: py-spark-train-basic

spec:

workspaces:

- name: source

description: Shared workspace for code + outputs

tasks:

# 1. Install Python requirements

- name: install-requirements

workspaces:

- name: src

workspace: source

taskSpec:

workspaces:

- name: src

steps:

- name: pip-install

image: python:3.11-slim

workingDir: $(workspaces.src.path)

script: |

#!/bin/sh

set -euo pipefail

echo "[install] $(date -Iseconds)"

if [ -f requirements.txt ]; then

python -V

pip install --upgrade pip

pip install -r requirements.txt

else

echo "No requirements.txt found, skipping install"

fi

# 2. Run Spark training

- name: train-model

runAfter: [install-requirements]

workspaces:

- name: src

workspace: source

taskSpec:

workspaces:

- name: src

steps:

- name: spark-train

image: bitnami/spark:3.3.1

workingDir: $(workspaces.src.path)

script: |

#!/bin/bash

set -euo pipefail

echo "[train] $(date -Iseconds)"

if [ ! -f model\_training.py ]; then

echo "ERROR: model\_training.py not found in $(pwd)" >&2

exit 1

fi

spark-submit \

--py-files ml\_stuff.zip \

model\_training.py

echo "Training finished."

**pipelinerun.yaml**

apiVersion: tekton.dev/v1

kind: PipelineRun

metadata:

generateName: py-spark-train-basic-run-

spec:

pipelineRef:

name: py-spark-train-basic

workspaces:

- name: source

volumeClaimTemplate:

spec:

accessModes: ["ReadWriteOnce"]

resources:

requests:

storage: 1Gi