AWS Certified Machine Learning Engineer - Associate Exam Guide

# Section 3.3 — Automate ML Workflows with CI/CD (In-Depth)

This section focuses on automation of ML workflows using CI/CD principles. It covers version control, pipeline orchestration, testing, deployment strategies, and AWS services for automation.

## 1. CI/CD Fundamentals

• Continuous Integration (CI) → automatically test and integrate code changes (unit, integration tests).

• Continuous Delivery/Deployment (CD) → automated deployment of models to production.

• Benefits: faster iteration, reproducibility, reduced manual errors.

## 2. Version Control

• Git basics: commits, branches, pull requests.

• GitFlow vs GitHub Flow for branching strategies.

• Repositories store training code, pipeline definitions, and model artifacts.

## 3. Pipeline Orchestration

• SageMaker Pipelines → end-to-end ML workflow automation (data prep → train → deploy).

• AWS Step Functions → orchestrate complex ML workflows with retries/error handling.

• Amazon EventBridge → trigger ML workflows on events (new data, commits).

## 4. AWS CI/CD Services

• AWS CodePipeline → orchestrates CI/CD pipelines.

• AWS CodeBuild → builds and tests code.

• AWS CodeDeploy → deploys models/services.

• AWS CodeCommit → managed Git repositories.

• Amazon ECR → stores container images for deployment.

## 5. Deployment Strategies

• Blue/Green Deployment → two environments (prod + staging), switch traffic.

• Canary Deployment → release to subset of users before full rollout.

• Linear Deployment → gradual traffic shift.

• Shadow Deployment → send prod traffic to new model without affecting results.

## 6. Testing & Validation

• Unit tests → validate code components.

• Integration tests → validate system interactions.

• End-to-end (E2E) tests → validate complete ML pipeline.

• Automated retraining triggers based on new data or drift detection.

## 7. Best Practices

• Automate retraining pipelines with SageMaker + EventBridge.

• Use Git for versioning code, models, and infrastructure definitions.

• Always include rollback strategies (blue/green, canary).

• Monitor CI/CD pipelines for failures and logs (CloudWatch).

## ✅ Exam Tips

• CodePipeline = orchestration, CodeBuild = testing/build, CodeDeploy = deployment.

• SageMaker Pipelines = ML-native orchestration tool.

• Canary & Shadow deployments = safe rollout strategies.

• Automated retraining often triggered by EventBridge.

• Always use Git-based workflows in exam scenarios.