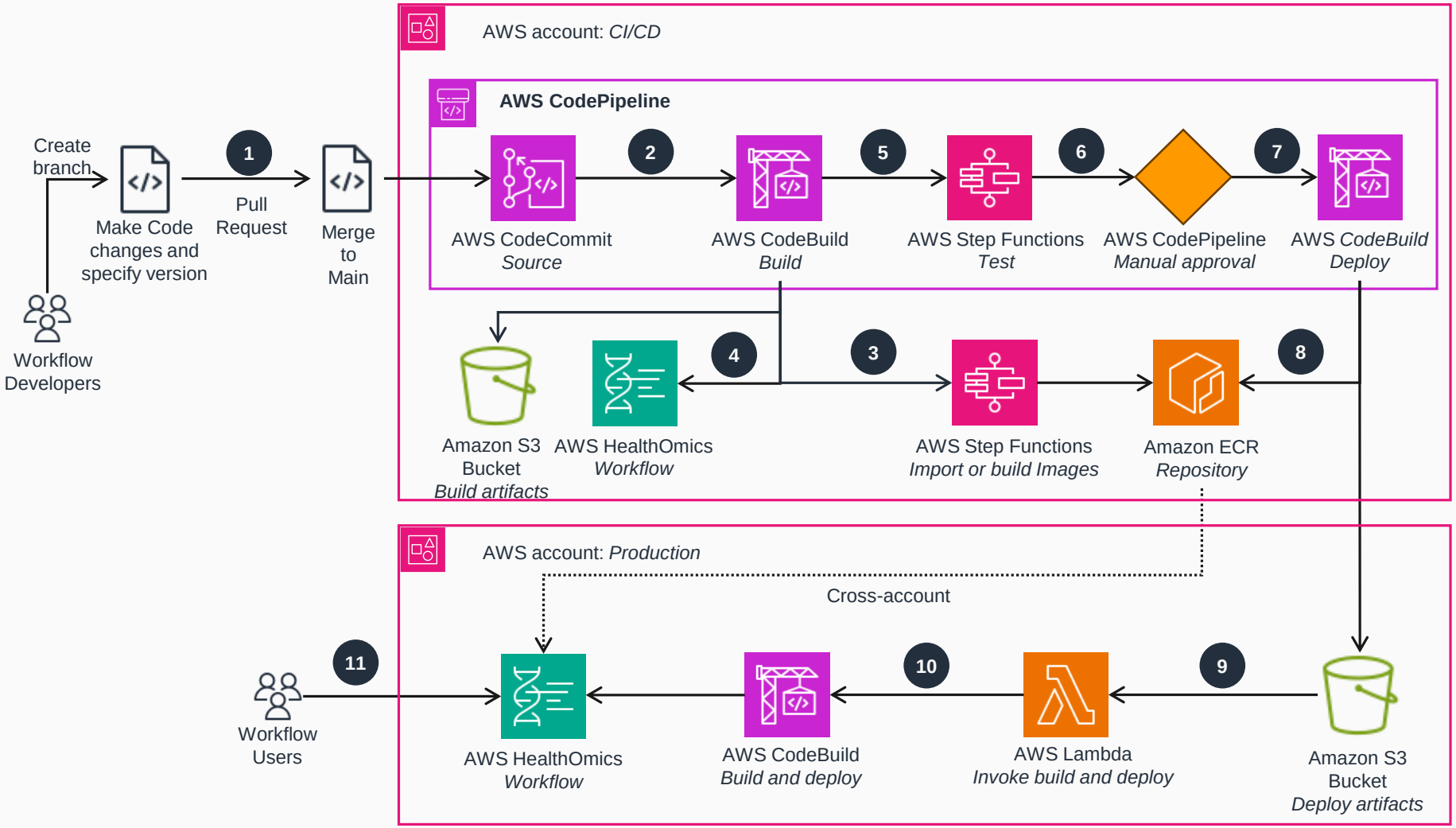
Guidance for Bioinformatics Workflow Development Using DevOps on AWS

This architecture diagram shows how to integrate DevOps best practices for bioinformatics workflow development on AWS.



1. Workflow developers create a branch and write code for a new workflow or make changes to an existing workflow repository within **AWS CodeCommit**. Developers can specify major and minor versions as part of semantic versioning. Since **CodeCommit** supports Git, developers use git operations to push code, submit a pull request, and merge to the 'main' branch. This runs the continuous integration and continuous delivery (CI/CD) pipeline configured with **AWS CodePipeline**.
2. Within **CodePipeline**, the build process starts using **AWS CodeBuild**, where it downloads the latest source code.
3. A pre-created **AWS Step Functions** state machine imports public Docker images or builds new Docker images and stores them within **Amazon Elastic Container Registry (Amazon ECR)** repositories.
4. The **CodeBuild** job prepares artifacts for the **AWS HealthOmics** workflow. It stores the artifacts in **Amazon Simple Storage Service (Amazon S3)** and creates a **HealthOmics** workflow. The workflow includes the user-defined semantic version and an auto-updated patch version.
5. After a successful build, **CodePipeline** runs a **Step Functions** state machine, which tests the **HealthOmics** workflow with some preconfigured test data and waits for completion.
6. On successful completion of the test workflow, a workflow administrator reviews the test workflow outputs in **Amazon S3** and, using **CodePipeline**, manually approves the deployment of the workflow to a production AWS account.
7. The approval action invokes a **CodeBuild** job that prepares the workflow artifacts and uploads them to the production account's **Amazon S3** bucket.
8. The **CodeBuild** job updates the **Amazon ECR** repository permissions for all applicable repositories to enable cross-account access for the production AWS account.
9. The upload of workflow artifacts to the production account's **Amazon S3** bucket runs an **AWS Lambda** function that checks for the necessary files and launches a **CodeBuild** job.
10. The **CodeBuild** job creates the workflow in **HealthOmics** using the artifacts in **Amazon S3**.
11. The **HealthOmics** workflow is now available for use in the production account.

Reviewed for technical accuracy May 2, 2024

AWS Reference Architecture