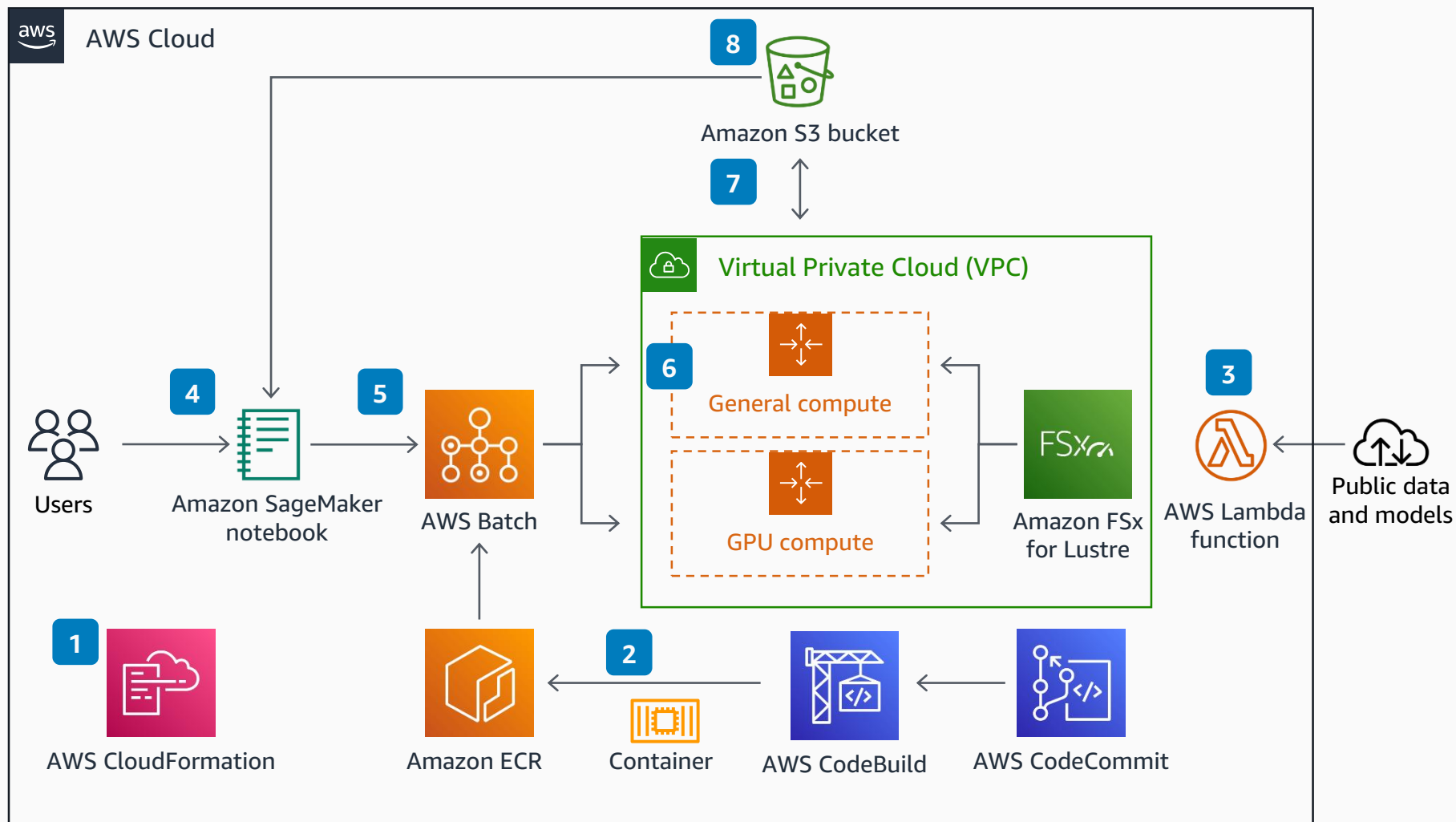


Guidance for Protein Folding on AWS

This architecture shows how to orchestrate protein folding and other analysis jobs on high performance computing (HPC) services, such as AWS Batch and Amazon FSx for Lustre.



- 1** **AWS CloudFormation** deploys the infrastructure in your AWS account.
- 2** **AWS CodeBuild** builds the containers necessary to run analysis algorithms, such as AlphaFold and OpenFold. All of the analysis algorithms are packaged as Docker containers and stored using **Amazon Elastic Container Registry (Amazon ECR)** in the deployment account. This helps ensure that all usage information remains private.
- 3** **AWS Lambda** triggers the download of model artifacts and reference data to an **Amazon FSx for Lustre** file system.
- 4** Define and submit analysis jobs from an **Amazon SageMaker** notebook instance or other Python environment.
- 5** **AWS Batch** manages job scheduling and orchestration.
- 6** Jobs run in general or accelerated compute environments based on the vCPU, memory, and GPU requirements.
- 7** Jobs write outputs and results to an encrypted **Amazon Simple Storage Service (Amazon S3)** bucket.
- 8** Users download job outputs to visualize the results or for downstream analysis.

