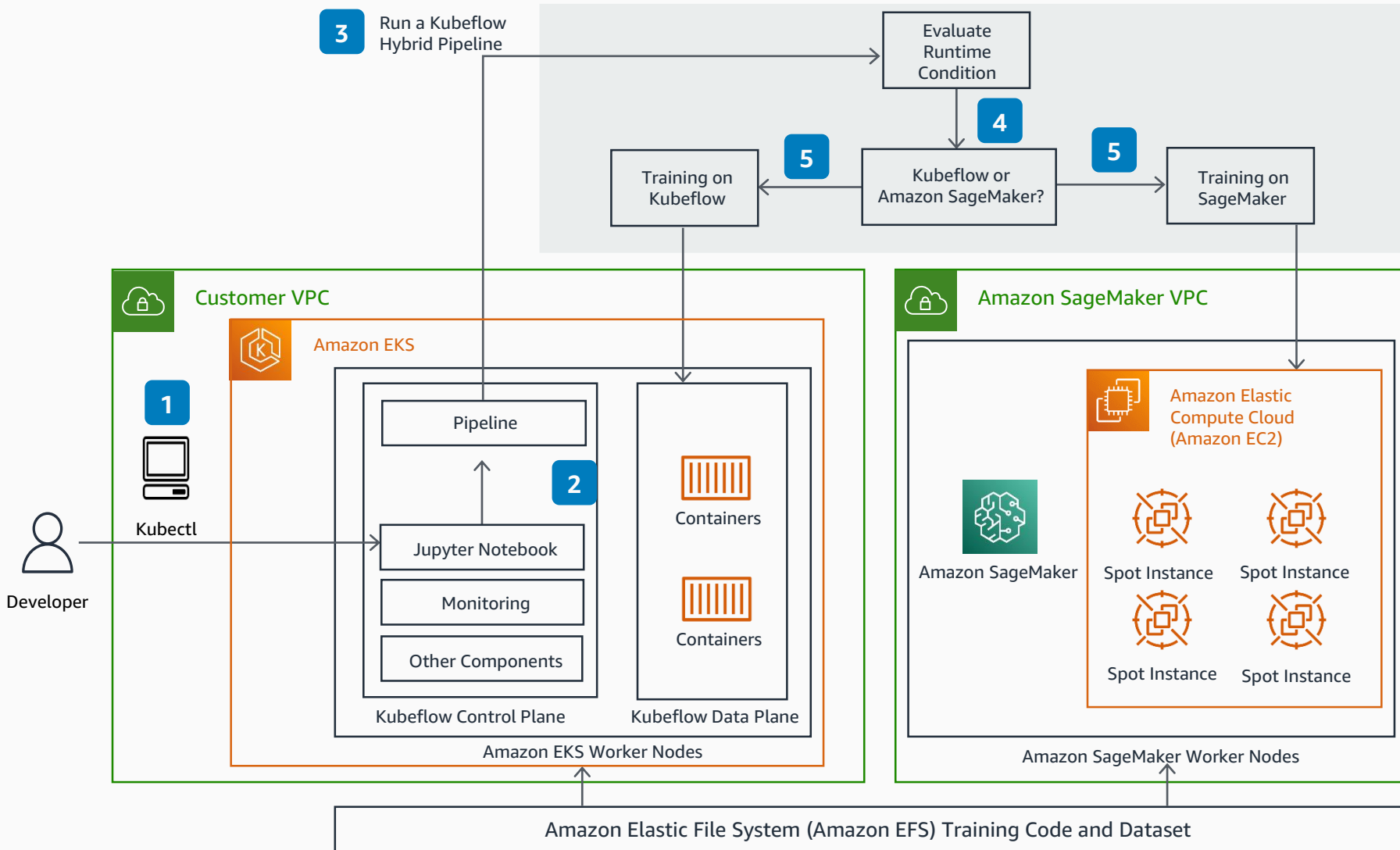


Guidance for Distributed Model Training on AWS

Building distributed training architectures using Kubeflow on AWS and Amazon SageMaker

A hybrid architecture enables workloads to transition between running on a self-managed Kubernetes or fully managed SageMaker infrastructure.



- 1 Deploy Kubeflow to **Amazon Elastic Kubernetes Service (Amazon EKS)** and access Jupyter Notebooks from the Kubeflow Central Dashboard. Kubernetes provides a command line tool (Kubectrl) for communicating with a Kubernetes cluster's control plane, using the Kubernetes API.
- 2 Use the Kubeflow Pipelines software development kit (SDK) to compile Python functions into workflow resources and to create Kubeflow pipelines.
- 3 Use the Kubeflow Pipelines SDK client to call the pipeline service endpoint and run the pipeline.
- 4 The pipeline evaluates the conditional runtime variables and decides between **Amazon SageMaker** or Kubernetes as the target run environment.
- 5 Use the Kubeflow PyTorch Operator to run distributed training on the Kubernetes cluster, or use the **SageMaker** component to submit the training on the **SageMaker** managed platform.

