

Lab: Deploying Specialized Infrastructure in AWS

Objective

Deploy a workload-optimized architecture using:

- EC2 with GPU acceleration
- FSx for Lustre

Part A: Console Steps

1. Launch a G5 instance with a deep learning AMI
2. Attach an FSx for Lustre file system
3. Install required ML libraries and test throughput

Part B: Terraform Steps

1. Deploy a G5 instance with EFA enabled
2. Mount FSx for Lustre via Terraform data source
3. Validate GPU access with `nvidia-smi`

Validation

- Verify FSx mount performance
- Run ML workload test to check GPU acceleration