## Prerequisites for Creating Databricks Workspace on AWS  
  
Before creating a workspace, ensure the following prerequisites are met:  
  
1. \*\*Databricks Account\*\*: You must have access to the Databricks Account Console.  
2. \*\*AWS Account\*\*: A valid AWS account with permission to create IAM roles, S3 buckets, and networking components.  
3. \*\*VPC\*\*: Your team (e.g., a cloud admin or network engineer) should create a compliant AWS VPC.  
4. \*\*Private Subnets\*\*: At least two private subnets in different availability zones.  
5. \*\*IAM Role\*\*: An IAM role with correct policies for S3 access and assume-role for Databricks.  
6. \*\*S3 Bucket\*\*: An S3 bucket for root storage of Databricks workspace.  
7. \*\*Security Group\*\*: A security group allowing the necessary inbound/outbound rules.  
8. \*\*Region Consistency\*\*: The VPC, subnets, S3 bucket, and Databricks workspace must be in the same AWS region.  
9. \*\*FedRAMP and HIPAA\*\*: Ensure the region supports FedRAMP Moderate and HIPAA if applicable.

This document provides a comprehensive guide to creating a production-ready Databricks workspace on AWS with HIPAA and FedRAMP Moderate compliance.

Production-Ready Databricks Workspace Setup with VPC, HIPAA & FedRAMP Moderate Compliance

## Step 1: Sign Up / Log in to Databricks Account

- Go to: https://databricks.com/  
- Sign in or create an account.  
- Navigate to your Databricks account console.  
  
Documentation: https://docs.databricks.com/administration-guide/account-settings-e2/account-create.html

## Step 2: Prepare Your AWS Environment

Ask your cloud administrator (or friend) to set up:  
- A VPC with:  
 - At least two private subnets in separate Availability Zones.  
 - Appropriate route tables, NAT gateway, and security groups.  
- An S3 bucket for root storage (e.g., `s3://databricks-prod-root/`).  
- IAM Role with policies allowing access to the S3 bucket and permissions to launch resources.  
  
Documentation:  
- VPC Requirements: https://docs.databricks.com/administration-guide/cloud-configurations/aws/customer-managed-vpc.html  
- IAM Role Setup: https://docs.databricks.com/administration-guide/cloud-configurations/aws/iam-role.html  
  
⚠️ Important: Ensure the following fields are consistent between the VPC and Databricks Workspace:  
- AWS Region (must be the same for all components)  
- VPC ID used  
- Private Subnet IDs (from the same VPC)  
- Security Group ID  
- IAM Role (cross-account trust enabled)  
- S3 Bucket location (same region as workspace)

## Step 3: Create a Databricks Workspace

1. Go to https://accounts.cloud.databricks.com/ > Workspaces > Create Workspace.  
2. Choose:  
 - Cloud: AWS  
 - Deployment: Customer-managed VPC  
 - Region: Same as the VPC region  
 - S3 Bucket: Your root S3 path  
 - IAM Role: Select previously configured IAM role  
3. Enable:  
 - HIPAA Compliance  
 - FedRAMP Moderate (on Enterprise or higher plans)  
4. Provide:  
 - VPC ID  
 - Private Subnet IDs  
 - Security Group ID  
  
Documentation: https://docs.databricks.com/administration-guide/account-settings-e2/workspaces.html

## Step 4: Configure Networking

Ensure secure connectivity via:  
- PrivateLink or VPC Peering  
- Subnet routing to allow access only to required services  
  
Documentation: https://docs.databricks.com/administration-guide/cloud-configurations/aws/network.html

## Step 5: Enable Account-Level Features

Post workspace creation:  
- Enable Unity Catalog for fine-grained access control  
- Setup audit logs to S3  
- Enable access control at account level  
  
Documentation:  
- Unity Catalog: https://docs.databricks.com/data-governance/unity-catalog/index.html  
- Audit Logging: https://docs.databricks.com/administration-guide/account-settings-e2/audit-logs.html

## Step 6: Secure the Workspace

- Use SCIM or SSO to onboard users  
- Assign role-based access controls (RBAC)  
- Configure IP Access Lists  
  
Documentation:  
- SCIM: https://docs.databricks.com/administration-guide/users-groups/scim.html  
- IP Access: https://docs.databricks.com/administration-guide/access-control/ip-access-list.html

## Step 7: Monitoring and Logging

- Integrate with AWS CloudWatch and CloudTrail  
- Enable workspace logging for clusters, jobs, and notebooks  
  
Documentation:  
- Job Monitoring: https://docs.databricks.com/workflows/jobs/jobs.html  
- Logging: https://docs.databricks.com/administration-guide/account-settings-e2/log-delivery.html

## Step 8: Validate Compliance Settings

For HIPAA:  
- Encrypt PHI in transit and at rest  
- Sign a BAA with Databricks  
  
For FedRAMP Moderate:  
- Enable audit logging  
- Use only FedRAMP-supported AWS regions  
  
Documentation:  
- HIPAA: https://docs.databricks.com/administration-guide/compliance/hipaa.html  
- FedRAMP: https://docs.databricks.com/administration-guide/compliance/fedramp.html

## Step 9: Operationalize Workspace

- Define cluster policies for cost and access control  
- Schedule jobs and alerts using Workflows  
- Use auto-scaling clusters  
  
Documentation:  
- Cluster Policies: https://docs.databricks.com/administration-guide/clusters/policies.html  
- Jobs: https://docs.databricks.com/workflows/jobs/jobs.html

## Step 10: Testing and Validation

- Test VPC connection with a notebook or job  
- Check user roles and permissions  
- Validate logging, cluster setup, and secure access