Databricks Accelerator

Rapid Data Lakehouse Foundation

Have you been searching for a way to seamlessly integrate Databricks to your AWS environment and empower your organisation to unlock the full potential of your data? The Databricks Accelerator is a solution designed by Databricks and AWS certified architects for turbocharging your Databricks implementation on AWS, delivering sustained reliability and simplified management.

Organizational enablers for operating Databricks Enterprise Platform

User Experience

Designed for easy usability - organisations can focus on delivering business value

Configuration

Adaptable to changing data and business needs through fully configurable templates

Scaling

Flexible, future proof architecture supports data growth and innovation

Reliability

Infrastructure as Code (IaC) and a fully automated deployment pipeline reduces effort to create, maintain, extend and operate Databricks & AWS resources

Observability

Amazon CloudWatch and Datadog Dashboards can be used for monitoring and visibility of all deployed resources

Security & Compliance

Well designed and optimised data and cloud infrastructure, ensuring account security and maximising data protection

Benefits of the Accelerator

The Databricks Accelerator offers a significant reduction of time to data readiness.

- **Observability**: Monitor resources with AWS tags, CloudWatch and Datadog
- **Configuration**: Customize via low-code according to business needs
- **Fully Automated**: Create all account resources and deploy via IaC and CI/CD pipelines
- **Risk Reduction**: Leverage best practices for data & cloud security
- **Integration**: Seamlessly integrates with other Thoughtworks Offerings

Databricks Accelerator

Rapid Data Lakehouse Foundation

Architectures may vary depending on custom configurations - The below sample reference architecture shows all entities and relationships between services that can be deployed on both Databricks and AWS.

Solution Architecture
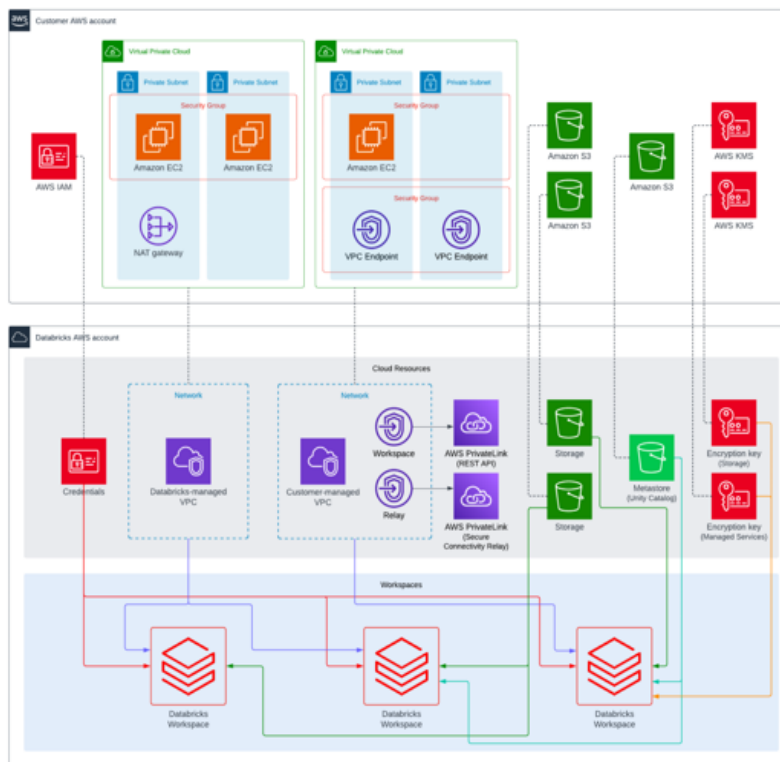
Technology Stack

- Databricks Enterpise Platform
- AWS Cloud Provider
- Terraform (Modules and Workspces)
- YAML
- Github and Github Actions

Prerequisites

- Databricks Enterpise Edition
- AWS Account
- Github Account

Configurables

Databricks
- Account and Workspaces
- User Management*
- Workflows and Jobs*

AWS
- Account and Regions
- VPC (Custom Databricks managed)
- Private Link & VPC Endpoints

When deploying Databricks workspaces via Terraform, the following resources are supported

Databricks Resources

- Workspace(s)
- Credential(s) (cross-account)
- Storage (DBFS)
- Network
- Encryption keys
- Metastore (Unity Catalog)

AWS Resources

- IAM roles and policies
- Amazon S3
- Amazon VPC
- Subnets, security groups, etc
- AWS KMS