

AWS Infra Engineer

JD-0055

Key Responsibilities

- Design, implement, and maintain AWS infrastructure including networking, compute, storage, and IAM to support production and non-production environments.
- Build and manage infrastructure-as-code (IaC) using tools such as Terraform or AWS CloudFormation for repeatable and automated provisioning.
- Architect and optimize environments to support agentic AI workloads, including model hosting, data pipelines, and integration with AI/ML services.
- Implement monitoring, logging, and alerting solutions (e.g., CloudWatch, CloudTrail, OpenSearch, Prometheus/Grafana) to ensure reliability and observability.
- Ensure security best practices across the AWS stack including VPC design, security groups, encryption, secrets management, and compliance controls.
- Collaborate with application, data, and AI teams to design CI/CD pipelines and deployment strategies for cloud-native and containerized workloads (with strong focus on Amazon EKS/ECS).
- Perform performance tuning, capacity planning, and cost optimization using AWS-native tools and governance practices.
- Troubleshoot complex infrastructure issues, perform root-cause analysis, and contribute to documentation and knowledge sharing.

Required Skills

- Strong hands-on experience with AWS core services (EC2, S3, VPC, IAM, RDS/Aurora, Lambda, Load Balancers, CloudWatch).
- Proven experience in infrastructure design, implementation, and operations for large-scale or business-critical systems.
- Practical expertise with Infrastructure-as-Code, with Terraform as a mandatory skill (CloudFormation experience is an added advantage).
- Experience building and managing infrastructure to support AI/ML or agentic AI applications (e.g., integration with SageMaker, Bedrock, custom model serving, or similar).
- Solid understanding of networking concepts (VPC, subnets, routing, security groups, VPN, Direct Connect, DNS, Transit Gateway, Inspection VPC).
- Strong background in Linux-based systems administration and shell scripting.
- Strong, hands-on experience with containerization and orchestration, specifically on Amazon ECS/EKS or Kubernetes, including designing and operating containerized solutions in production.
- Knowledge of CI/CD tools and pipelines (e.g., GitLab CI, GitHub Actions, Jenkins, AWS CodePipeline/CodeBuild).
- Good understanding of cloud security best practices, identity and access management, and compliance considerations.
- Strong troubleshooting skills and ability to work in a fast-paced, collaborative environment.

Good to Have Skills

- Experience with AWS AI/ML services (SageMaker, Bedrock, Lex, Kendra, Comprehend) or other cloud AI platforms.

- Familiarity with agentic AI frameworks, orchestration tools, or LLM application frameworks.
- Experience with serverless architectures (Lambda, API Gateway, Step Functions).
- Knowledge of data engineering components on AWS (Glue, Kinesis, Redshift, EMR, Athena).
- Exposure to observability stacks and APM tools (Datadog, New Relic, Dynatrace, or similar).
- Experience implementing cloud governance, tagging strategies, and FinOps practices.
- Knowledge of security frameworks and standards (CIS Benchmarks, ISO 27001, SOC 2) and related tooling.
- Background in Python or other scripting languages for automation and tooling development.

Experience Required

- 6 to 8 years of relevant experience in AWS infrastructure engineering or cloud platform engineering roles.