Deployment Architecture

Deployment Architecture

> **Project:** {{PROJECT_NAME}} **Version:** {{VERSION}} **Date:** {{DATE}}
> **Author:** {{AUTHOR}} **Status:** Draft | In Review | Approved **Reviewers:** {{REVIEWERS}}

Document History

Version	Date	Author	Changes
0.1	{{DATE}}	{{AUTHOR}}	Initial draft

---

1. Overview

System: {{PROJECT_NAME}} **Cloud Provider:** {{CLOUD_PROVIDER}} **Provider Rationale:** {{RATIONALE}} **Architecture Pattern:** {{PATTERN}}

---

2. Infrastructure Topology

```
graph TB
    subgraph Internet
        USER[End Users]
        CDN[CDN / CloudFront]
    end

    subgraph Public Subnet
        ALB[Application Load Balancer]
        BASTION[Bastion Host]
```

```
    end

    subgraph Private Subnet - App
        APP1[App Server 1]
        APP2[App Server 2]
    end

    subgraph Private Subnet - Data
        DB_PRIMARY[(Primary DB)]
        DB_REPLICA[(Read Replica)]
        CACHE[Redis Cache]
    end

    subgraph Isolated Subnet
        SECRETS[Secrets Manager]
        BACKUP[Backup Storage]
    end

    USER --> CDN
    CDN --> ALB
    ALB --> APP1
    ALB --> APP2
    APP1 --> DB_PRIMARY
    APP2 --> DB_PRIMARY
    APP1 --> CACHE
    DB_PRIMARY --> DB_REPLICA
    APP1 --> SECRETS
```

---

3. Networking Architecture

3.1 VPC / VNET Design

Network	CIDR	Purpose
VPC / VNET	{{CIDR_VPC}}	Main network boundary
Public Subnet A	{{CIDR_PUB_A}}	Load balancers, NAT gateways
Public Subnet B	{{CIDR_PUB_B}}	Load balancers, NAT gateways (AZ-B)

Network	CIDR	Purpose
Private Subnet A	{{CIDR_PRIV_A}}	Application servers
Private Subnet B	{{CIDR_PRIV_B}}	Application servers (AZ-B)
Isolated Subnet A	{{CIDR_ISO_A}}	Databases, secrets
Isolated Subnet B	{{CIDR_ISO_B}}	Databases, secrets (AZ-B)

3.2 Load Balancer Configuration

Parameter	Value
Type	{{LB_TYPE}}
Protocol	HTTPS (TLS 1.2+)
SSL Termination	At load balancer
Health Check Path	{{HEALTH_CHECK_PATH}}
Health Check Interval	{{INTERVAL}}s
Unhealthy Threshold	{{THRESHOLD}} consecutive failures
Idle Timeout	{{TIMEOUT}}s
Stickiness	{{STICKINESS}}

3.3 DNS Architecture

Record	Type	Value	TTL
{{DOMAIN}}	A / ALIAS	Load Balancer	{{TTL}}
api.{{DOMAIN}}	CNAME	API Load Balancer	{{TTL}}
cdn.{{DOMAIN}}	CNAME	CDN Distribution	{{TTL}}

DNS Provider: {{DNS_PROVIDER}} **Failover Strategy:** {{FAILOVER_STRATEGY}}

3.4 CDN Configuration

Parameter	Value
Provider	{{CDN_PROVIDER}}
Origin	{{CDN_ORIGIN}}
Cache Behaviors	Static assets: 1yr, API: no-cache, HTML: 5min
HTTPS Only	Yes

Parameter	Value
WAF Integration	{{WAF_INTEGRATION}}

---

4. Compute

4.1 Container Orchestration

Platform: {{ORCHESTRATION}}

Component	Configuration	Notes
Cluster	{{CLUSTER_SPEC}}	
Node Groups	{{NODE_GROUPS}}	
Min Nodes	{{MIN_NODES}}	
Max Nodes	{{MAX_NODES}}	
Node Size	{{NODE_SIZE}}	
Container Registry	{{REGISTRY}}	

4.2 Serverless Functions

Function	Trigger	Memory	Timeout	Purpose
{{FUNCTION_1}}	{{TRIGGER}}	{{MEMORY}}MB	{{TIMEOUT}}s	{{PURPOSE}}

4.3 Instance Sizing & Auto-Scaling

Service	Instance Type	Min	Max	Scale Trigger
{{SERVICE}}	{{INSTANCE}}	{{MIN}}	{{MAX}}	CPU > {{CPU}}% for {{DURATION}}min

Scale-Out Policy: {{SCALE_OUT}} **Scale-In Policy:** {{SCALE_IN}} **Scale-In Cooldown:** {{COOLDOWN}}min

---

5. Storage

5.1 Database Hosting

Database	Engine	Version	Hosting	Instance	Storage	HA
{{DB_NAME}}	{{ENGINE}}	{{VERSION}}	{{HOSTING}}	{{INSTANCE}}	{{STORAGE}} GB	{{HA}}

Connection Pooling: {{POOL_TOOL}} **Max Connections:** {{MAX_CONN}} **Connection String:** Stored in {{SECRET_LOCATION}} (never hardcoded)

5.2 Object Storage

Bucket / Container	Purpose	Access	Lifecycle	Encryption
{{BUCKET_NAME}}	{{PURPOSE}}	{{ACCESS}}	{{LIFECYCLE}}	AES-256

5.3 File Storage

Storage	Type	Mount Point	Purpose	Size
{{STORAGE_NAME}}	{{TYPE}}	{{MOUNT}}	{{PURPOSE}}	{{SIZE}}GB

---

6. Security

6.1 Network Security Groups / Firewall Rules

Security Group	Direction	Port	Protocol	Source / Destination	Purpose
sg-alb	Inbound	443	TCP	0.0.0.0/0	HTTPS from internet
sg-alb	Outbound	{{APP_PORT}}	TCP	sg-app	Forward to app
sg-app	Inbound	{{APP_PORT}}	TCP	sg-alb	From load balancer
sg-app	Outbound	{{DB_PORT}}	TCP	sg-db	Database access
sg-db	Inbound	{{DB_PORT}}	TCP	sg-app	From application only

6.2 WAF Configuration

WAF Provider: {{WAF_PROVIDER}}

Rule Group	Purpose	Action
AWSManagedRulesCommonRuleSet	OWASP Top 10	Block
AWSManagedRulesSQLiRuleSet	SQL injection	Block
AWSManagedRulesKnownBadInputsRuleSet	Known bad inputs	Block
Rate limiting	{{RATE_LIMIT}} req/5min per IP	Count → Block

6.3 Secrets Management

Secret Store: {{SECRET_STORE}}

Secret	Rotation Schedule	Access
Database credentials	90 days	App role only
API keys (third-party)	On compromise	App role only
TLS certificates	60 days before expiry	Deploy role only
JWT signing key	365 days	Auth service only

6.4 IAM Roles & Policies

Role	Trusted By	Key Permissions	Purpose
{{APP_ROLE}}	EC2 / ECS Task	SecretsManager:GetSecret, S3:GetObject	Application runtime
{{DEPLOY_ROLE}}	CI/CD	ECR:PushImage, ECS:UpdateService	Deployments
{{BACKUP_ROLE}}	Lambda / Cron	RDS:CreateSnapshot, S3:PutObject	Backups

---

7. Cost Estimation

Component	Service	Spec	Est. Monthly Cost
Compute	{{SERVICE}}	{{SPEC}}	${{COST}}
Database	{{SERVICE}}	{{SPEC}}	${{COST}}
Load Balancer	{{SERVICE}}	{{SPEC}}	${{COST}}

Component	Service	Spec	Est. Monthly Cost
CDN	{{SERVICE}}	{{TRAFFIC}}GB transfer	${{COST}}
Storage	{{SERVICE}}	{{CAPACITY}}GB	${{COST}}
Monitoring	{{SERVICE}}	{{METRICS}} metrics	${{COST}}
Total			**${{TOTAL}}**

Cost Optimization Notes:

-

---

8. High Availability Design

Component	HA Strategy	Failover Time	Notes
Application	Multi-AZ, N+1 instances	Immediate (ELB health check)	
Database	Multi-AZ with auto-failover	60-120 seconds	DNS propagation
Cache	Cluster mode / Replication	30 seconds	Redis Sentinel
CDN	Global edge network	Transparent	Provider HA

RTO Target: {{RTO}} minutes **RPO Target:** {{RPO}} minutes

---

9. Multi-Region Considerations

Current: {{REGION_STRATEGY}} **Primary Region:** {{PRIMARY_REGION}} **Secondary Region:** {{SECONDARY_REGION}}

Rationale: {{MULTI_REGION_RATIONALE}}

Data Replication: {{REPLICATION_STRATEGY}} **Failover Procedure:** See disaster-recovery-plan.md

---

10. Related Documents

- CI/CD Pipeline
- Environment Configuration

- Infrastructure as Code
- Monitoring & Observability
- Disaster Recovery Plan

Approval

Role	Name	Date	Signature
Author			
Reviewer			
Approver			

Revision #7
Created 2026-02-23 12:05:33 UTC by John
Updated 2026-05-25 07:33:42 UTC by John