Cloud Infrastructure & Architecture Standard (Multi-Region)

Document ID: IE-CAS-2025.01
Version: 1.0
Effective Date: 2025-11-01
Department: IT & Engineering - Infrastructure

1. Cloud Provider and Strategy

NexaCore utilizes a **multi-cloud strategy** with a primary focus on [Major Cloud Provider] for its AI/ML training workloads and primary SaaS hosting. All infrastructure is managed using **Infrastructure as Code (IaC)** via Terraform/Pulumi.

2. Global Architecture Model

Our core platform utilizes a three-tier architecture, implemented across multiple geographic regions to serve our global client base and meet data localization requirements:

Region Type	Purpose	Key Services Hosted	Required Compliance
Primary Region (Dubai/DIFC)	Core R&D, primary API endpoint, central user authentication, core database.	ML Training Cluster, Kubernetes Control Plane, Main Application API Gateway.	UAE PDPL, Financial Services (DIFC).
MENA PoP (KSA/Egypt)	Regional Point-of-Presence (PoP) for data localization and low-latency client access.	Regional Caching, Load Balancing, Data Shards (isolated customer data).	KSA/Egypt Data Sovereignty Laws.
DR Region (Europe)	Disaster Recovery (DR) and cold backup site.	Database backups, static asset storage, standby compute cluster.	GDPR (as fallback).

3. Core Principles

- **Scalability:** All services must be deployed using containerization (Kubernetes) and auto-scaling groups. Fixed-size resources are prohibited for production workloads.
- **Cost Management:** Utilize reserved instances and spot instances for non-critical, scalable compute (e.g., specific batch processing). Cost monitoring is mandatory via FinOps tools.
- **Network Isolation:** Implement strict **Zero Trust** principles. All VPCs must be logically segmented, with no default routes between environments (Dev, Staging, Production).
- **Monitoring:** Mandatory integration with the central **System Monitoring and Alerting Standard**.