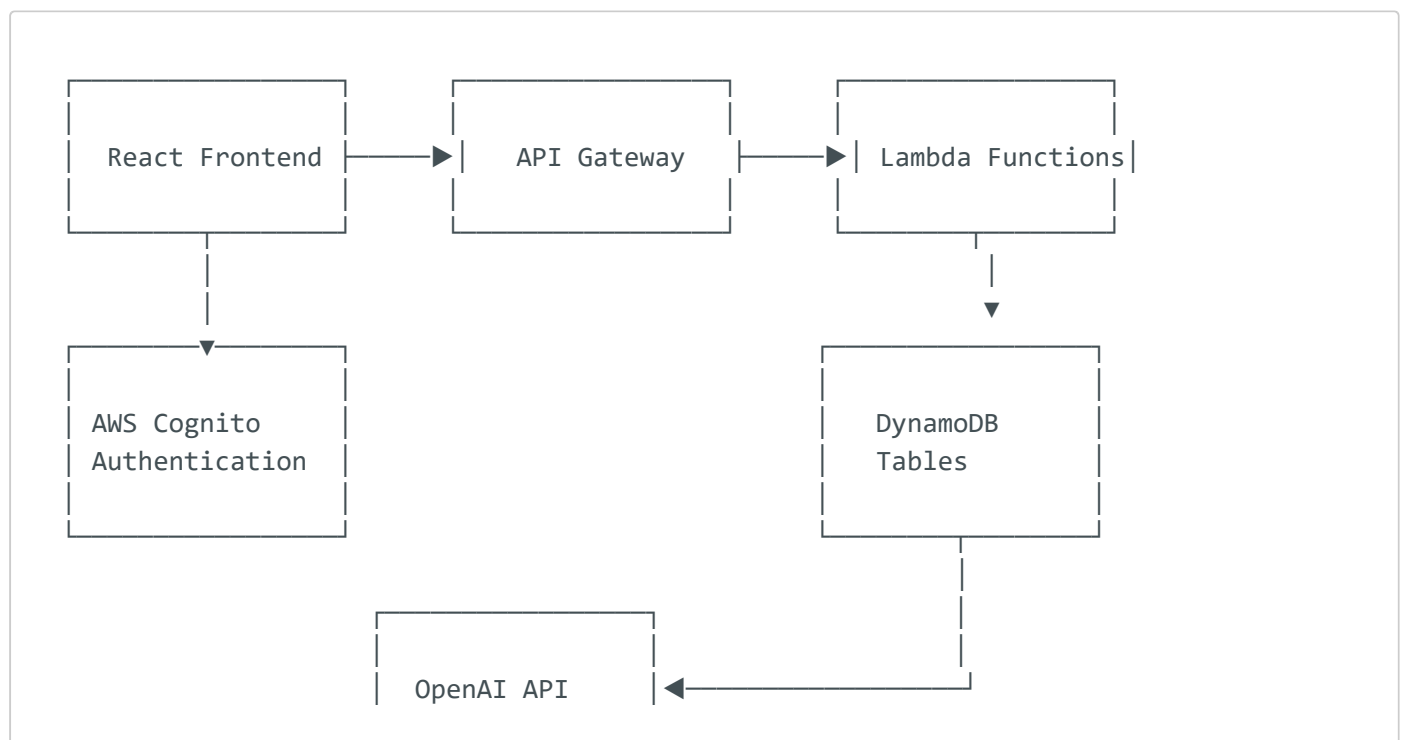


- [Document Processing Accelerator: System Architecture](#)
 - [System Overview](#)
 - [Core Components](#)
 - [Frontend Layer](#)
 - [API Layer](#)
 - [Business Logic Layer](#)
 - [Data Storage Layer](#)
 - [AI Integration Layer](#)
 - [Security Layer](#)
 - [Data Flow](#)
 - [Deployment Architecture](#)
 - [Scalability Considerations](#)
 - [Regional Deployment](#)

Document Processing Accelerator: System Architecture

System Overview

The Document Processing Accelerator is a serverless solution that leverages AWS services and AI to automate document processing workflows. This document outlines the high-level architecture, components, and interactions within the system.



Core Components

Frontend Layer

- **React Application:** TypeScript-based SPA for document upload, viewing, and management
- **AWS Amplify:** Integration with AWS services
- **Authentication:** Cognito integration for secure access

API Layer

- **API Gateway:** RESTful API endpoints with the following resources:
 - `/documents`: Document management operations
 - `/documents/{id}/process`: Trigger document processing
 - `/documents/upload-url`: Generate pre-signed S3 URLs
 - `/swagger`: Swagger documentation endpoints

Business Logic Layer

- **Lambda Functions:** Serverless compute for document processing:
 - Document metadata management
 - Upload URL generation
 - Document processing orchestration
 - AI integration

Data Storage Layer

- **DynamoDB:** NoSQL database for document metadata storage
- **S3 Buckets:** Object storage for document files
 - Raw document storage
 - Processed results storage

AI Integration Layer

- **OpenAI API:** AI service for document understanding and extraction
 - Text extraction from documents
 - Named entity recognition
 - Document classification
 - Data structuring

Security Layer

- **Cognito:** User authentication and authorization
- **IAM Roles:** Fine-grained access control
- **S3 Pre-signed URLs:** Secure document upload

Data Flow

1. Document Upload Flow

User → Frontend → API Gateway → Lambda → S3 (via pre-signed URL)

2. Document Processing Flow

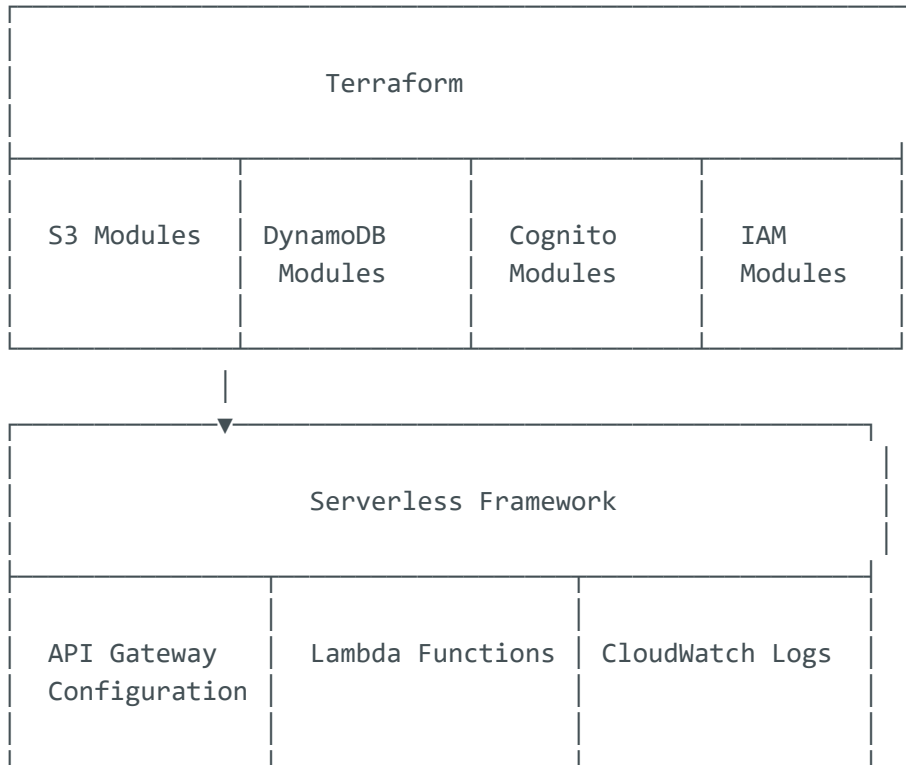
S3 → Lambda → OpenAI API → Lambda → DynamoDB

3. Document Retrieval Flow

User → Frontend → API Gateway → Lambda → DynamoDB/S3 → Frontend

Deployment Architecture

The system is deployed using Infrastructure as Code (IaC) principles:



The architecture leverages two complementary IaC tools:

- **Terraform:** Manages core AWS infrastructure components
- **Serverless Framework:** Deploys and manages the Lambda functions and API Gateway

Scalability Considerations

The serverless architecture provides inherent scalability benefits:

- **Lambda Auto-scaling:** Functions scale automatically based on workload
- **DynamoDB On-demand Capacity:** Database scales with usage patterns
- **S3 Infinite Scale:** Document storage grows without limits
- **API Gateway Throttling:** Configurable rate limiting for API protection

Regional Deployment

The system is designed to be deployed to a single AWS region with multi-region capability through configuration changes.

Default region: `us-east-1`