

AWS PDF Translator

Serverless Document Translation Application

📍 Region: **eu-west-1 (Ireland)**

🏗️ Architecture: **Serverless**

📅 Date: **January 2026**

React

Python 3.11

AWS Lambda

Amazon S3

DynamoDB

API Gateway

CloudFront

Cognito

CodePipeline

1. Objective

The primary objective of this project is to design and implement a fully serverless, cloud-native PDF translation application that enables users to automatically translate PDF documents from one language to another using Amazon Web Services (AWS).

- **Automated Translation:** Develop a system that automatically extracts text from uploaded PDF documents, detects the source language, and translates the content into the user's desired target language using neural machine translation.
- **Serverless Architecture:** Build a cost-effective, scalable solution using AWS serverless services (Lambda, API Gateway, S3, DynamoDB) that requires zero server management and scales automatically based on demand.
- **Secure User Authentication:** Implement robust user authentication and authorization using Amazon Cognito to ensure that each user's documents and translations remain private and secure.
- **CI/CD Automation:** Establish a continuous integration and continuous deployment pipeline using AWS CodePipeline and CodeBuild integrated with GitHub, enabling automated deployments with every code change.
- **Global Accessibility:** Deploy the frontend application through Amazon CloudFront CDN to provide fast, low-latency access to users worldwide.

2. Introduction

In today's globalized world, the ability to quickly and accurately translate documents across languages has become essential for businesses, researchers, educators, and individuals. Traditional translation methods are often time-consuming, expensive, and require manual intervention. This project addresses these challenges by leveraging the power of cloud computing and artificial intelligence to deliver an automated, on-demand PDF translation service.

The **AWS PDF Translator** is a modern web application built entirely on serverless architecture. Users can upload PDF documents through an intuitive React-based web interface, select their desired target language, and receive professionally translated documents within minutes. The system employs Amazon Translate, a neural machine translation service that delivers high-quality, natural-sounding translations across 75+ languages.

The application architecture follows AWS best practices for serverless design, incorporating multiple services that work together seamlessly: Amazon S3 provides secure object storage for uploaded and translated documents; AWS Lambda executes the business logic without provisioned servers; Amazon API Gateway exposes RESTful endpoints; Amazon DynamoDB stores job metadata with single-digit millisecond latency; Amazon Cognito handles user authentication; and Amazon CloudFront delivers the frontend globally with edge caching.

To support modern DevOps practices, the project includes a fully automated CI/CD pipeline. Developers can push code changes to GitHub, triggering AWS CodePipeline to automatically build, test, and deploy updates to both the frontend application and Lambda functions, reducing deployment time from hours to minutes.

3. Conclusion

The AWS PDF Translator project successfully demonstrates the implementation of a production-ready, serverless document translation application on Amazon Web Services. By combining multiple AWS services into a cohesive architecture, the project achieves its core objectives of automation, scalability, security, and cost-effectiveness.

Key Achievements:

- **Fully Functional Translation System:** The application successfully extracts text from PDF documents, automatically detects source languages using Amazon Comprehend, translates content using Amazon Translate's neural engine, and delivers translated documents through secure presigned URLs.
- **Scalable Serverless Architecture:** The implementation proves that complex document processing workflows can be built entirely on serverless services, automatically scaling from zero to thousands of concurrent translations without infrastructure management.
- **Secure Multi-User Platform:** Amazon Cognito integration provides enterprise-grade authentication, ensuring users can only access their own documents. All data is encrypted at rest and in transit.
- **Automated DevOps Pipeline:** The GitHub-integrated CI/CD pipeline enables rapid, reliable deployments within 3-5 minutes, supporting agile development practices.
- **Cost-Efficient Operation:** The pay-per-use model ensures costs correlate directly with usage. For low-volume usage (under 50 translations/month), total cost remains under \$20.

Future Enhancements

The modular architecture supports future enhancements including support for additional document formats (Word, PowerPoint), batch translation capabilities, translation memory for improved consistency, custom terminology support, and integration with enterprise content management systems.

This project serves as a reference architecture for organizations looking to build document processing applications on AWS, demonstrating best practices in serverless design, security implementation, and DevOps automation.