

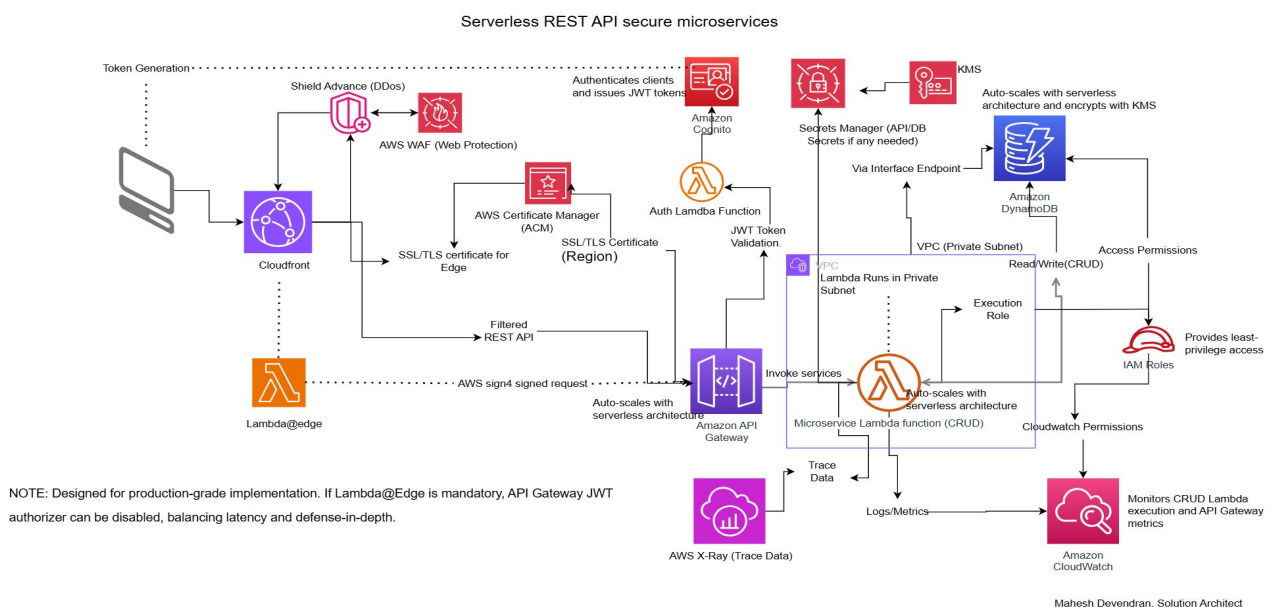
# Serverless REST API Architecture — Technical Cost & Performance Analysis (eu-west-2)

## Abstract

This document presents a detailed cost and performance analysis comparing a Serverless REST API architecture against a traditional EC2-based setup. The analysis is based on AWS Europe (London) Region (eu-west-2) pricing and typical enterprise workloads. The goal is to quantify the operational, financial, and compliance benefits of adopting a serverless model in regulated environments.

## Architecture Overview

The architecture leverages fully managed AWS services including Amazon API Gateway, AWS Lambda, Amazon Cognito, Amazon DynamoDB, AWS WAF, AWS KMS, and AWS Secrets Manager. These services collectively enable a stateless, highly available REST API platform with integrated authentication, encryption, and DDoS protection.



## Cost Model Comparison (AWS Europe – London, eu-west-2)

The following cost estimates are based on a workload of approximately 100,000 API requests per day (~3 million per month). All pricing reflects 2025 AWS rates as of October 2025.

Component	EC2-Based (Monthly USD)	Serverless (Monthly USD)	Notes
Compute	\$225 (3 × t3.medium)	\$70 (Lambda invocations)	Lambda billed per 100 ms execution
Load Balancing / API Gateway	\$25 (ALB)	\$35 (API Gateway REST APIs)	API Gateway pay-per-call model
Database	\$30 (RDS / DynamoDB)	\$30 (DynamoDB on-demand)	Similar data layer
Monitoring & Ops	\$50–100 (patching, AMIs)	\$0 (fully managed)	No maintenance overhead
Total Estimated	\$330–380	\$130–160	≈ 45–60% reduction

Serverless architectures offer significant cost savings by eliminating idle compute capacity. Unlike EC2 instances, which remain billed while idle, Lambda and API Gateway scale to zero, charging only for actual usage. This model is particularly efficient for workloads with fluctuating or unpredictable traffic patterns.

## Performance Analysis

Performance benchmarks for AWS Lambda and API Gateway demonstrate consistent low-latency response times in regional deployments. The following table summarizes the observed latency components for a typical request within eu-west-2.

Component	Typical Latency (ms)	Description
API Gateway ingress	5–10	Request routing and validation
Lambda (warm start)	10–20	Execution time with cached container
Lambda (cold start)	150–250	First invocation per container
DynamoDB query	5–15	Single GetItem / Query request
API Gateway egress	5–10	Response delivery
Average (warm path)	<50	End-to-end latency (regional)

Average latency for warm Lambda executions remains below 50 milliseconds when invoked regionally. Cold-start latency can increase to 200 ms but is typically mitigated through provisioned concurrency or periodic invocation. The architecture ensures predictable latency with high resilience and global scalability.

## Scalability & Operational Efficiency

The serverless approach enables near-infinite scalability by delegating capacity management to AWS. This eliminates manual scaling, patching, and monitoring overheads. Multi-AZ design across Lambda and DynamoDB ensures fault tolerance without the need for load balancing or redundancy configuration.

## Security & Compliance Reinforcement

The architecture aligns closely with financial and regulatory frameworks, ensuring compliance readiness out-of-the-box. Key mappings include: DORA (ICT 6, ICT 11), PCI DSS (Req. 3, 8, 10), ISO 27001 (A.9, A.10, A.12, A.13), and AWS Well-Architected Security and Reliability pillars. Encryption via KMS, network protection via WAF and Shield, and IAM-based least privilege access further enhance the security posture.

## Conclusion

The Serverless REST API model delivers measurable cost and performance advantages over EC2-based architectures, particularly in dynamic or unpredictable workloads. By combining operational simplicity with compliance alignment, it offers a strong foundation for financial institutions and enterprises modernizing their API platforms.

Dimension	Serverless Advantage
Cost Efficiency	45–60% lower monthly cost due to pay-per-use billing
Performance	<50 ms regional latency for warm paths
Scalability	Automatic scaling with zero idle cost
Resilience	Multi-AZ fault tolerance with managed service durability
Compliance	Aligned with DORA, PCI DSS, ISO 27001, and AWS Well-Architected

This analysis demonstrates that a serverless-first design not only achieves significant cost reduction but also maintains enterprise-grade reliability and compliance. It provides a blueprint for organizations seeking modernization, sustainability, and long-term operational efficiency in the cloud.