MCIT
Modeling, control & information technologies

# *Improving Performance and Reliability of Electronic Document Issuance Systems Using Caching and Asynchronous Processing*

https://doi.org/10.31713/MCIT.2025.056

Vitalii Koverdiuk
National University of Water and Environmental Engineering
Rivne, Ukraine
v.v.koverdiuk@nuwm.edu.ua

Oksana Ostapchuk
National University of Water and Environmental Engineering
Rivne, Ukraine
o.p.ostapchuk@nuwm.edu.ua

*Abstract*—**This document focuses on the optimization of electronic document issuance systems that operate under high user load. As the volume of digital services and e-government platforms increases, the need for faster, more reliable, and scalable document generation becomes crucial. The study investigates key bottlenecks in document processing pipelines and proposes a hybrid approach based on server-side caching and asynchronous task execution. The proposed solution significantly reduces response times and improves the overall throughput of document generation systems.**

**Keywords — electronic document issuance; caching; asynchronous processing; scalability; system reliability;**

## I. Introduction

Electronic document issuance systems (EDIS) have become a cornerstone of modern e-government, enterprise automation, and financial ecosystems. They are responsible for generating, validating, signing, and distributing digital documents — such as certificates, invoices, licenses, and official notices — in real time. With the steady increase in demand for remote administrative and legal services, these systems are now required to handle thousands of concurrent requests per second while maintaining security and data integrity.

However, traditional monolithic or synchronous system designs face scalability barriers. As the number of users grows, the load on backend databases, signing services, and file storage becomes critical. Performance issues such as delayed responses, queue congestion, and timeouts lead to a poor user experience and service downtime. Moreover, repetitive document generation — often for identical templates or metadata — causes redundant computations and unnecessary I/O operations, further degrading performance.

In addition, regulatory compliance and cybersecurity requirements impose additional complexity: each document must be digitally signed, timestamped, and archived, all within short timeframes. To ensure consistent performance under these constraints, a new approach to system design is required — one that integrates parallelism, caching, and asynchronous task orchestration into a unified architecture.

This research aims to explore and evaluate methods for improving the performance and reliability of EDIS, focusing on asynchronous processing and caching techniques that enable responsive, fault-tolerant, and scalable operation in high-load environments [1–3].

## II. PROBLEM STATEMENT

Existing EDIS platforms commonly use a synchronous processing model: a user request triggers sequential operations such as template retrieval, data binding, document rendering, and digital signing. Each of these steps is typically I/O-bound and blocking, resulting in resource underutilization and slow response times under load. Key challenges include:

- high latency due to sequential task execution;
- excessive database load caused by repeated queries;
- redundant template compilation and metadata retrieval;
- limited scalability of monolithic architectures;
- lack of asynchronous coordination between modules.

Therefore, there is a need for an architectural redesign that separates request handling from processing logic, introduces concurrency, and minimizes repetitive data operations while maintaining document authenticity and compliance [4].

## III. OVERVIEW OF EXISTING SOLUTIONS

Recent research and engineering practice highlight several effective approaches to improving the performance and scalability of electronic document issuance systems (EDIS). The most common techniques include load balancing, microservice architecture, database optimization, and caching.

Load balancing and horizontal scaling are widely used to distribute workload across multiple computing nodes, ensuring stable performance under high traffic. Cloud-native platforms such as Kubernetes or Docker Swarm automate this process, dynamically allocating resources according to system load.

Another major improvement is the shift from monolithic to microservice and event-driven architectures. By decoupling document generation, signing, and delivery processes, EDIS can process requests concurrently and avoid single points of failure. Asynchronous communication through message queues (RabbitMQ, Kafka, Azure Service Bus) enables non-blocking workflows and increases reliability.

Database performance is also a critical factor. Query caching, indexing, and the use of in-memory systems such as Redis significantly reduce latency and prevent overload. Caching of templates and metadata shortens rendering time and lowers CPU utilization, especially in systems generating large volumes of similar documents.

Finally, monitoring and auto-scaling tools like Prometheus and Grafana help maintain system stability under fluctuating loads. They allow dynamic resource allocation based on real-time performance metrics.

While these methods individually improve certain components, few implementations integrate them into a single coherent architecture. A unified model combining asynchronous processing, caching, and automatic scalability provides a more efficient foundation for high-load EDIS environments..

## IV. PROPOSED SOLUTION

The proposed architecture integrates asynchronous processing and multi-level caching to enhance performance, scalability, and reliability of electronic document issuance systems. This section outlines the main concepts and benefits of the approach:

- asynchronous task execution: each user request is placed in a distributed queue and processed by background workers. This allows the main API to respond immediately while document generation continues independently;
- event-driven workflow: the system operates in a non-blocking mode, executing multiple generation and signing tasks concurrently without waiting for previous ones to complete;
- multi-level caching: frequently accessed data and templates are stored in several layers — data cache for metadata, template cache for precompiled layouts, and result cache for recently issued documents — which minimizes repetitive computations;
- integrity and authenticity: each cached document is protected with a digital signature, timestamp, and cryptographic hash, ensuring that its content remains verifiable and unchanged;
- dynamic scaling and monitoring: resource allocation automatically adjusts to workload intensity, while monitoring tools (Prometheus, Grafana) detect bottlenecks and maintain system stability;

Experimental results under simulated load (10,000 concurrent users) revealed that the proposed solution significantly improved system performance. The average response time decreased by up to 80%, while backend throughput increased by 40–50%. CPU utilization remained stable even under peak traffic conditions, and no document integrity violations were detected during asynchronous issuance, confirming both the reliability and efficiency of the architecture.

## CONSCLUSION

The research demonstrates that the integration of asynchronous processing and multi-level caching provides an efficient and reliable solution for optimizing high-load electronic document issuance systems. The asynchronous architecture allows multiple generation and signing processes to run concurrently, maintaining responsiveness even under peak demand. Meanwhile, caching reduces redundant computations and database queries, improving overall throughput and ensuring balanced use of system resources.

In addition, the use of digital signatures, timestamps, and integrity checks preserves the authenticity of issued documents, while dynamic scaling and continuous monitoring maintain stable operation across varying workloads. Altogether, the proposed approach enhances performance, reliability, and scalability, offering a sustainable foundation for future cloud-based e-government and enterprise digital platforms..

## REFERENCES

[1] M. Kleppmann, “Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems” O’Reilly Media, 2022.

[2] S. Newman, “Building Microservices: Designing Fine-Grained Systems 2ed.” O’Reilly Media, 2021.

[3] S. Cleary, “Concurrency in C# Cookbook,” O’Reilly Media, 2019.

[4] M. Abbott, M. Fisher, “The Art of Scalability: Scalable Web Architecture, Processes, and Organizations for the Modern Enterprise” Addison-Wesley Professional, 2015.

[5] P. Membrey, E. Plugge, D. Hows, “Practical Load Balancing: Ride the Performance Tiger” Apress, 2012

[6] M. Richards, N. Ford, “Fundamentals of Software Architecture: An Engineering Approach” O’Reilly Media, 2020.

[7] S. Cleary, “Concurrency in C# Cookbook” O'Reilly Media, October 2019.