

Nikkei revamped search for over 210 million records on Elastic Cloud, the foundation for NIKKEI KAI GenAI

Nikkei, publisher of Japan's leading economic news and database services, migrated its 210 million-document archive to Elasticsearch on Elastic Cloud, cut infrastructure costs 22.5%, and used RAG with hybrid vector search to launch its NIKKEI KAI generative AI service in under a year.

22.5%

Infrastructure cost reduction

210M+

Documents searchable, back to 1876

<1 yr

NIKKEI KAI GenAI, concept to production

~200

On-prem servers before the Elastic migration

The challenge

- ✗ Nikkei's 210 million-document archive ran on-premises across roughly 200 servers, so each new media partnership meant adding hardware and operational load.
- ✗ Database management was outsourced, so every requirements change meant rewriting specs and paying for external development, costing both money and time.
- ✗ Version upgrades were difficult, and a small in-house team could not easily operate a large-scale search system on its own.
- ✗ Even on self-managed Elasticsearch on AWS, infrastructure overhead stayed heavy and unexplained errors were slow to diagnose.

The solution

- ✓ Nikkei migrated the article database and search engine to Elasticsearch, then to the managed Elastic Cloud service in 2024.
- ✓ Fault tolerance and replicas give automatic node recovery and high availability, run and upgraded by a small in-house team.
- ✓ Managed Elastic Cloud removed day-to-day infrastructure work and gave the team Elastic responders to lean on for issues.
- ✓ Hybrid retrieval combining vector and full-text search, plus RAG through Elasticsearch, powers the NIKKEI KAI GenAI service.

How it works

1. Ingest & index

Over 210 million records indexed in Elasticsearch

2. Decompose query

User question split into sub-queries

3. Hybrid retrieve

Vector plus full-text search, then RAG

4. Generate answer

GenAI returns a direct, zero-click answer

Before

- On-prem search across ~200 servers; scaling meant buying and managing more hardware
- Outsourced database changes: rewrite specs, pay for external dev, then wait
- Heavy operational overhead; unexplained errors were slow to diagnose

After

- Managed Elastic Cloud runs the 210 million-document archive with a small team
- 22.5% lower infrastructure cost, with in-house upgrades and scaling
- RAG and hybrid search launched the NIKKEI KAI GenAI service in under a year

"By adopting the managed service Elastic Cloud, we **no longer need to manage infrastructure**, allowing us to **focus on service development**."

— Yasufumi Mizoguchi, Service Infrastructure First Group, Technical Strategy Unit, Nikkei Inc.

Why it matters beyond one archive

Nikkei's search revamp was the groundwork that made generative AI viable. A scalable, managed retrieval layer let the team ship NIKKEI KAI in under a year. Next, it plans article chunking and a sparse vector model like ELSER to compress vector data across over 200 million documents, extending intuitive, beyond-keyword search to the entire archive.