



Data Modernization

Global Financial Institution

How We Re-Engineered a 40+ Billion Document Search Platform to Cut Costs and Boost Performance

A global financial institution, operating at significant scale across multiple markets. With tens of thousands of employees and vast volumes of transactional data, performance and efficiency are critical to daily operations.

At the heart of one of their key services was a document search platform. This system allowed teams to query business critical metadata across billions of records and over time, the system had grown complex.

It supported highly flexible, open-ended searches across more than 1,200 metadata fields. While powerful, this flexibility came at a cost. The platform relied on dozens of indexes to keep up with demand, driving up storage, memory usage, and IOPS.

Performance was starting to suffer and costs were rising; the system was also becoming harder to maintain.

The client needed a more efficient way forward and asked gravity9 to help redesign their approach.



Review of the Challenges

This was not a simple optimisation; the platform handled over 40 billion documents, with a highly dynamic structure. Each document could contain a different combination of metadata fields, making indexing and querying difficult to standardize.

The existing setup relied on:

- Over 27 indexes to support flexible search
- A custom built query layer using JSON-based DSL
- Dynamic query construction using Strin based assembly that was prone to errors

The client's goal was clear: move towards a more efficient search model using MongoDB Atlas Search but getting there was complex.

Key challenges included:

- Translating custom query logic into a new search framework
- Reworking the data model to support the new approach
- Handling migration at massive scale
- Ensuring no disruption to existing services

Even small changes carried risk at this level of volume.

Our Solution

We redesigned the core of the system without breaking what already worked. Instead of replacing everything, we extended the existing library to support both old and new search methods. This allowed the client to transition safely, without disruption.

At the center of the solution was a new query conversion layer; this layer translates incoming JSON based queries into the correct format for MongoDB Atlas Search. Unlike the previous approach, it builds queries in a structured and reliable way, reducing errors and improving maintainability.

We also introduced:

- A routing layer to decide when to use legacy or new search logic
- A cleaner abstraction layer to remove hard-coded business logic
- A shared library to eliminate duplication across teams

Alongside this, we developed migration scripts capable of handling tens of billions of records. These were designed to run in parallel, recover from failure, and support large scale data transformation.

The result was a system that is more robust, more efficient, and ready to scale.

Utilized Technology Stack:

Java (backend):

MongoDB Atlas

AWS

Gatling (for load simulation)

JMeter (for load testing)

NodeJS (for data migration scripts)



Our Approach

We began with a deep dive into the existing system this included analyzing the data model, reviewing the query library, and studying real production query patterns. This helped us identify incorrect assumptions early and avoid costly rework later.

We worked closely with both the client and MongoDB engineers to refine the solution. Where gaps or risks were identified, we raised them early and adapted the approach.

Testing was critical; we built an automated test suite that:

- Converted JSON DSL queries into both old and new formats
- Executed them side by side
- Compared results for consistency

This allowed rapid validation and helped uncover issues such as unsupported data types early in the process.

The approach ensured accuracy, reduced risk, and kept the project moving forward despite complexity.

Subsequent Outcomes

The redesigned solution delivered clear benefits with the platform now positioned to:

- Reduce reliance on large numbers of indexes
- Lower storage, memory, and IOPS costs
- Improve query performance and response times
- Simplify future development and maintenance

By supporting both legacy and new search methods, the client can transition at their own pace, without disrupting operations. The migration framework also provides a scalable path for handling extremely large datasets, with built-in resilience and recovery.

While the wider programme was paused due to internal changes, the work delivered created a strong technical foundation for future adoption.

Client Feedback: Going Beyond the Brief to Deliver Real Value

The feedback received was highly positive, they recognized not only the technical delivery, but the team's proactive approach. In particular, they valued the early identification of data model issues and solution gaps which was well beyond the original scope of rewriting the query library.

The team was also praised for clear communication and strong collaboration.

The platform is now positioned to reduce costs, improve performance, and scale with confidence.

Visit our Insights page for more articles about emerging technology trends, the Finance Industry, interviews, and more!