MarkLogic®

The Haufe Group Implements a Cloud-based MarkLogic® Content Hub to Manage Complex Content

The Haufe Group Experiences Improved Search Results Thanks to Modern Database Technology

HAUFE.Group

····························

CUSTOMER
The Haufe Group

INDUSTRY
Media & Publishing

FEATURES

CUSTOMER USE CASE
Central Content Hub
Data Integration
360 Degree View of Data
Deletion of Content Duplication

BENEFITS
Automation of Required
Document Updates

Accelerated and Streamlined
Cycle of Content Production to
Delivery

Significant Improvement of the
Search Results

Company Overview

The Haufe Group is one of the leading providers of digital workplace solutions and services in Germany. Customers rely on the company's expertise, which is based on huge amounts of data. The data must be skillfully managed and linked so that users can quickly get the most comprehensive information possible about the range of specialist literature as well as additional options that furthers education based on their searches. To allow users to access the full range of products and services and achieve accurate search results of their keyword searches, the Haufe Group uses MarkLogic's database in its Editorial Department.

The Haufe Group has evolved from a successful publishing business into a specialist for digital and web-based products and services. Licensees access specialized information from various work areas via paid portals like Human Resources, Tax, Law, Occupational Safety, Civil Service and Compliance. In addition, the Haufe Group operates an academy with a broad range of seminars for specialists and executives.

··

Challenges

Before the introduction of the new database technology, the company had distributed its data across different data silos, each created for different applications. A linking of the data and thus an overall view of all content was hardly possible. Due to the silos, users would receive results of their keyword searches which only partly reflected the total availability of specialist information and topical cross-references were largely missing. The scattering losses for the user were significant, and Haufe wanted to fix this issue.

The original system was based on the open-source platform Apache Solr™ and relational databases. It stored different types of data such as book metadata, widgets, book covers, product information, and the actual content in different data silos. Queries were made using various search engines and tools.

The goal was to build a central content hub, which eliminates these deficits and provides users with much better search results. In total, around 1.5 million documents (mostly in XML format) had to be integrated into a new database. The Haufe Group found that importing and converting files is time-consuming and complex in relational databases.

The Solution

After an evaluation phase and the development of a proof of concept (PoC), a NoSQL database based on a flexible data model was most suited for the task. Specifically, the Haufe Group chose MarkLogic running on Microsoft® Azure®.

In the first step, the concept of the Content Hub was created and then a rough data model was developed, which transformed different content into a uniform format. The aim was to make the automation of frequently required document updates as easy as possible while at the same time significantly improving the search results for users.

The quick and uncomplicated updating of documents is of central importance to the Haufe Group because the company's know-how is based on very detailed and varied content creation for the special departments. "Our editorial team has always been very content-oriented. This way of working had to be transferred to digital document management as well," said Alexander Bieber, Project Manager Content Hub at the Haufe Group.

The central element of the architecture is a layer of services that is exposed to consumers through an API gateway. In detail, these are services for ingesting the content (processing content in the database) and for searching and analyzing the content. The other supporting element is MarkLogic, which supports the services layer and acts as a search engine and central document repository. MarkLogic provides flexibility, a schema-less data model and high scalability.

Why MarkLogic

The introduction of MarkLogic has greatly improved queries and thus the quality of search results. The result lists are sorted according to various topics (Controlling, Finance, Law, Social Welfare, etc.), and the user also receives information about the type and number of documents found, such as news stories, comments, work aids, downloads, and more. This gives users the opportunity to further refine searches and improve results.

The linked search has a particularly positive effect. For example, users searching for tax-relevant topics in the real estate sector may also discover relevant documents from the tax department that are also available. This was not possible with the old system due to data storage in different data silos. With a click on the respective article, the system also provides an overview of the most widely read articles on the subject, available downloads, seminar offers and topic-relevant specialist magazines. "The MarkLogic database provides a complete 360-degree view of all existing documents. The added value for the user lies in the granular search options, which ensures very good results," said Bieber.

The database at the Haufe Group currently processes an average of 20 to 30 queries per second, sometimes reaching as fast as 80 queries per second. This puts the company in an excellent position, allowing users to access a much larger pool of data with the new system than was possible before the introduction of MarkLogic.