indium
Make Technology Work
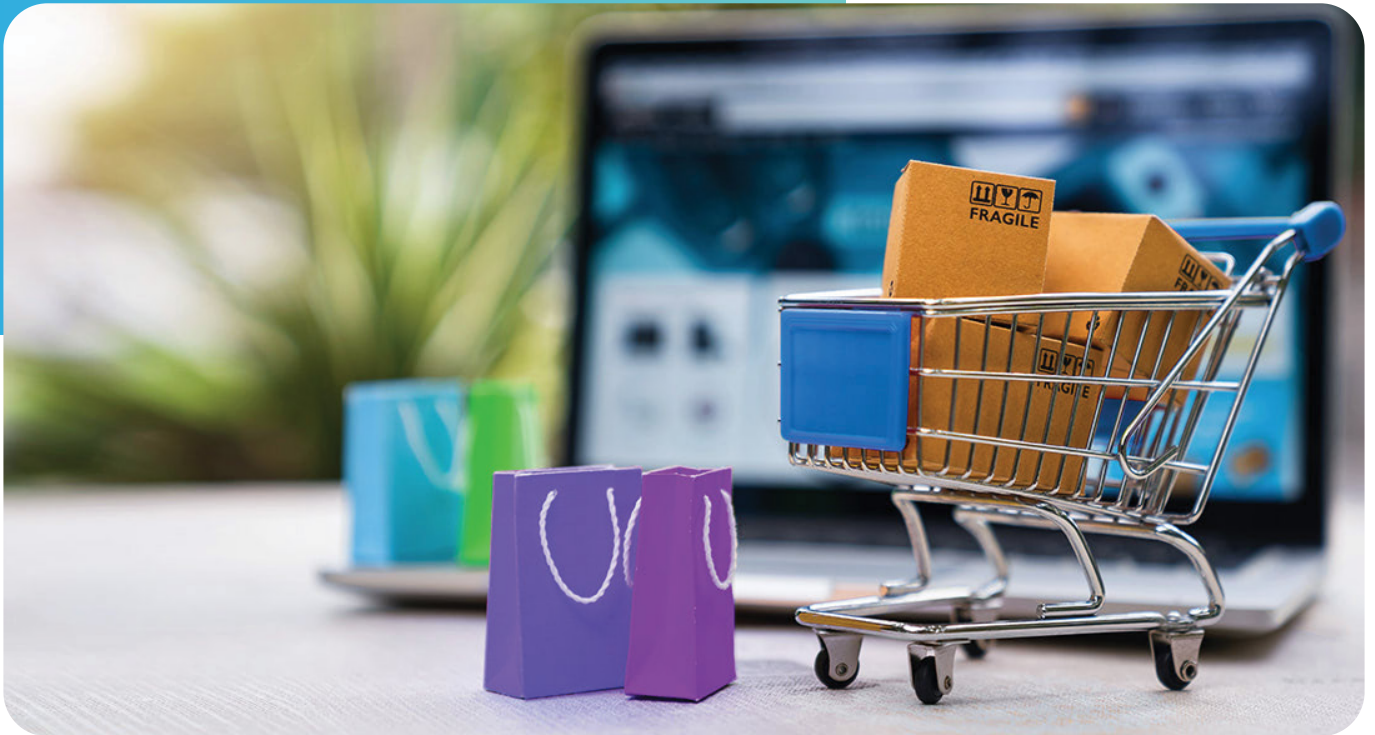

DATA INDEXING USING ELASTICSEARCH ENGINE FOR AN E-COMMERCE PRICE AGGREGATOR

PROJECT OVERVIEW

As the retail industry increasingly shifts online, the growing number of online platforms, including e-commerce sites, second-hand trading boards, retailer-owned sites, and online catalogues, has made it challenging for consumers to compare prices across all options and find the best deals. In response, Indium Software developed a solution that provides real-time visibility of products and dynamically changing prices across multiple sites. The solution utilizes an efficient data handling system that processes real-time data requests and responses with minimal delay. Additionally, it effectively crawls numerous e-commerce sites to extract and present the most relevant information in the correct taxonomy.

SOLUTION DELIVERED

Data Engineering

CLIENT DOMAIN

eCommerce

KEY HIGHLIGHTS

- Implemented efficient data storage using batch processing, reducing processing time from 25 hours to 7 hours.

- Achieved near real-time data refresh and response time of less than 1 second, providing instant price information without any latency.

ABOUT CLIENT

The client operates an internet-based price comparison website, providing consumers with quick comparison information for e-commerce products to help them find the best deals. Their platform aggregates data from over 1500+ retailer websites, enabling users to have real-time visibility of products and their dynamically changing prices. To meet this challenge, the client required an efficient data handling system capable of processing real-time data requests and responses.

BUSINESS REQUIREMENTS

- Categorizing and redefining the gathered data from multiple eCommerce sources into a standardized format before storing it, to address the product taxonomy challenge.

- Achieving a response time of 1 second or less for processing data refresh requests, in order to meet the end user's expectation of receiving price updates within 2 seconds.

- Addressing the issue of huge data storage loads resulting from collecting data from over 1500+ retailer websites. The current single line indexing approach for storing data is causing inefficiencies, with the average data storage process taking approximately 25 hours.

SOLUTION HIGHLIGHTS

The solution highlights are as follows:

- Indium Software implemented a powerful Elasticsearch engine in a Big Data environment for efficient data indexing.

- A machine learning algorithm using Google Taxonomy was implemented to standardize the product data, addressing the product taxonomy challenge.

- The Big Data ecosystem was utilized to process billions of data points, resulting in a response time of less than 1 second for any number of query requests.

- The solution leveraged the ETL (Extract, Transform, Load) batch processing capabilities of Elasticsearch, enabling the processing of 10GB of data in approximately 1 second. This helped solve the single line indexing problem, improving data storage efficiency.

BUSINESS IMPACT

- The implementation of efficient data storage using batch processing reduced the process time from 25 hours to 7 hours, improving data storage efficiency.

- The solution demonstrated powerful data load carrying capacity, allowing for the collection and storage of data from over 1500 retailer websites.

- The data refresh and response time were optimized to near 1 second, providing real-time price information to users without any latency. This enhanced the user experience and allowed for quick price comparisons.

TECH STACK

elasticsearch

Java

python

RDS

aws

ABOUT INDIUM

Indium Software is a fast-growing Digital Engineering company, focused on building modern solutions across Applications, Data, and Gaming for its clients. With deep expertise in next-gen offerings combining data and applications, Indium offers a wide range of services including Product Engineering, Low-Code development, Data Engineering, Ai/ML, Digital Assurance, and end-to-end Gaming services.

indium
Make Technology Work

USA	INDIA	UK	SINGAPORE
Cupertino | Princeton	Chennai | Bengaluru | Mumbai | Hyderabad	London	Singapore
Toll-free: +1-888-207-5969	Toll-free: 1800-123-1191	Ph: +44 1420 300014	Ph: +65 6812 7888

www.indiumsoftware.com

For Sales Inquiries
sales@indiumsoftware.com

For General Inquiries
info@indiumsoftware.com