

TECHNICAL VALIDATION

# Maximizing Search Application Performance with Elasticsearch

By Alex Arcilla, Senior Validation Analyst  
Enterprise Strategy Group

June 2023

# Contents

Introduction.....	3
Background.....	3
Elasticsearch .....	4
Enterprise Strategy Group Technical Validation.....	5
Simple Query and Sort.....	6
Sorting with Timestamp, Keyword, and Numeric Data .....	6
Date Histogram Aggregation Using Timestamped Data .....	7
Terms Query and Aggregation .....	8
Ranges.....	9
Resource Utilization.....	10
Conclusion.....	11

## Introduction

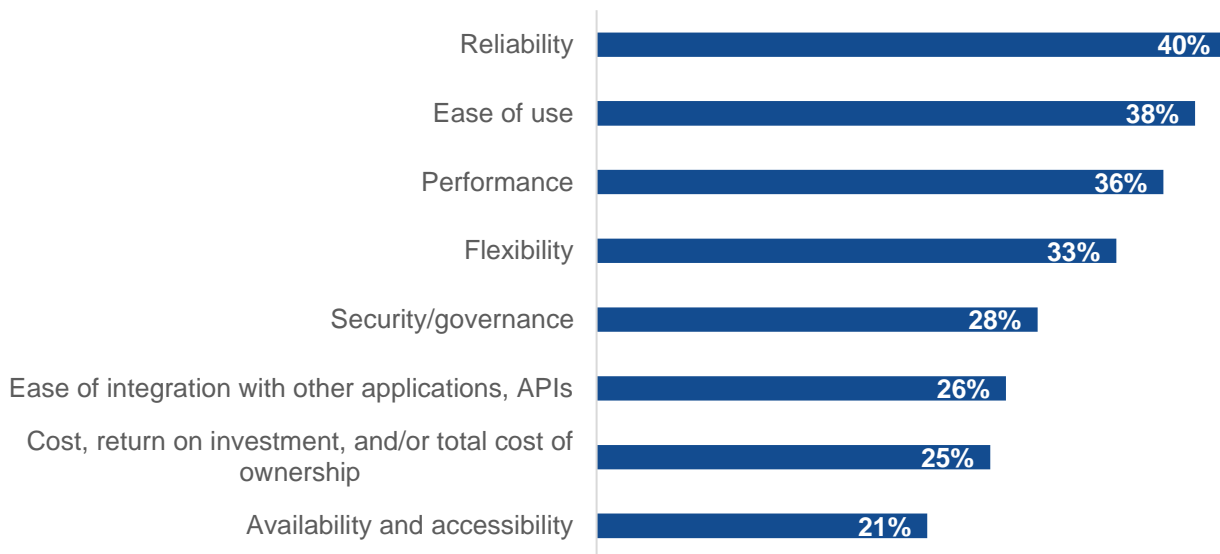
This Technical Validation from TechTarget's Enterprise Strategy Group documents our evaluation of Elasticsearch. We compared Elasticsearch with a search platform from a leading vendor using test results that characterize performance and scalability such as search queries and resource utilization.

## Background

When choosing an effective search platform, the speed and relevance of search results is critical to meet business needs. Internal search applications, designed to help employees find specific tools or files, need to deliver relevant results quickly to improve collaboration and knowledge sharing. With e-commerce sites, organizations want to drive conversions to maximize sales. For customer support applications, organizations must ensure that employees can save time in finding the most relevant information, helping not only to save costs but also retain customers. Some businesses are built entirely on embedded search applications for cases such as matching people with available car rides. In fact, when it comes to implementing technologies for supporting data initiatives, Enterprise Strategy Group research uncovered that 36% of respondents cited performance as one of the top five most important attributes that they look for (see Figure 1).<sup>1</sup>

**Figure 1. Top Eight Most Important Capabilities When Supporting Data Initiatives**

**When looking to implement technologies to support data initiatives within your organization, what are the most important capabilities/attributes? (Percent of respondents, N=403, five responses accepted)**



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

When evaluating search platforms, organizations need to consider many factors to ensure that the technology will sufficiently address their business needs. Amongst the most critical factors to consider are:

<sup>1</sup> Source: Enterprise Strategy Group Research Report, [The State of DataOps](#), August 2022.

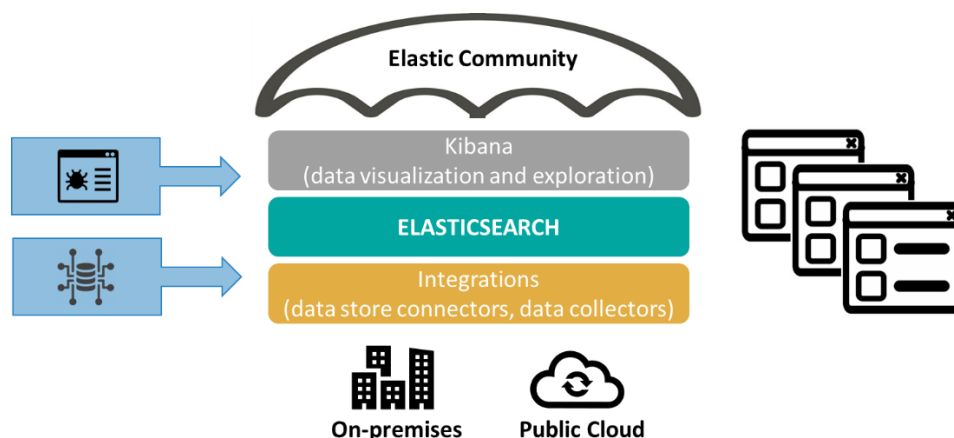
- **Performance:** Will the search platform discover the relevant data and produce search results in a short period of time, regardless of the size of the underlying data set?
- **Scalability:** Can the search platform scale to handle the continuing growth and distributed nature of data—structured and unstructured—without negatively impacting access and retrieval of the most relevant data needed and, subsequently, performance?
- **Resource optimization:** Does the search platform efficiently use the underlying infrastructure (e.g., storage memory) so that costs are minimized, whether running on premises or in the public cloud?
- **Flexibility:** Can organizations operate the search platform on-premises or in the public cloud as business needs dictate?
- **Support:** Will developers and end users receive the support they need in terms of continuous feature integration and issue resolution so that search applications are optimized and available?

## Elasticsearch

Elasticsearch is designed to help organizations build highly performant search-powered applications. Built on Apache Lucene, Elasticsearch is a distributed analytics engine for all types of data, including textual, numerical, geospatial, structured, and unstructured. The search platform offers numerous ways to ingest data from multiple and disparate sources. Once ingested, the platform searches this data to discover, analyze, and index the most relevant results for any query.

Figure 2 illustrates how components of the platform work to maximize performance, scale, and optimize search results. Connectors ingest data from data stores, located on premises or in the public cloud, or via Elastic's built-in web crawler. To optimize search results, Elasticsearch offers multiple capabilities, powered by machine learning, to increase search result relevancy, including vector and hybrid search (for semantic searches), neurolinguistic programming (for personalization), and adaptive relevance (for curation suggestions). To monitor and optimize performance and availability, organizations can leverage the platform's Kibana support to build dashboards and visualizations for monitoring native search logs and metrics.

**Figure 2.** Elasticsearch



*Source: Enterprise Strategy Group, a division of TechTarget, Inc.*

Since Elasticsearch offers multiple built-in capabilities to manage and monitor search performance, organizations can decrease tool sprawl, thus reducing operational expenses. Tool sprawl can be further reduced when

Elasticsearch is used for security and observability use cases. Organizations can then achieve a better return on investment when unifying search, security, and observability operations using Elasticsearch.

Elasticsearch can be run on-premises and in both hybrid cloud and public cloud environments. With this flexibility, organizations can easily adapt to changing data requirements, such as where data is stored and retrieved for search queries. And organizations can expect the same benefits of performance, scalability, and resource optimization.

To support implementation of Elasticsearch, organizations can leverage over 300 built-in data integrations covering commonly used enterprise applications and data sources from which customers ingest data into the platform. More importantly, developers can enlist the support of the Elasticsearch community that possesses skill sets around Elasticsearch and Kibana. Since the community encourages open collaboration and communication, developers can improve their own projects by using the guidance and code contributions continuously posted by members.

## Enterprise Strategy Group Technical Validation

Enterprise Strategy Group reviewed the results of tests that compared the performance, scalability, and resource optimization of Elasticsearch (v8.7) to an alternative search engine (comparable in feature support) offered by a leading cloud service provider (CSP). (The CSP will be referred to as Vendor X throughout the report.) The testing method employs standard methodologies, openly available data sets, and industry-standard tools and practices. This is critical to note as these tests were designed to be repeatable.

The test bed consisted of two Kubernetes clusters running on Google Cloud Platform (GCP), with each cluster running Elasticsearch or the competitive offering. Clusters consisted of five E2 machine types, each with eight CPUs, 32GB of RAM, and 300GB of disk storage, deployed in the US-Central-1a availability zone via Terraform. By using this testbed, the infrastructure can be completely abstracted so that testing can easily be replicated on other public clouds such as Amazon Web Services, Microsoft Azure, or IBM Cloud.

To simulate real-world workloads, such as indexing and searching, tests used reproducible data generated by the Elastic Corpus Generator, an open source tool. Datasets up to 1TB were created to run queries and estimate the performance results highlighted in this report.

Over a 19-day period, multiple runs were conducted on each search platform. Each run repeated specific queries 100 times. If a query failed during a test run, all results of that run were discarded.

To compare performance between the search platforms, we evaluated test results based on the 90th percentile (or confidence level),<sup>2</sup> a statistical measure that is the value **below which 90% of the observations fall**. We used this metric because commonly used summary statistics (e.g., the mean or average) can be adversely influenced by outliers. Using the confidence level minimizes how outliers influence how the data is interpreted and is more appropriate in comparing performance between Elasticsearch and Vendor X's offering.<sup>3</sup>

Finally, both Elasticsearch and the alternative offering were configured with the same settings and parameters to ensure that the performance differences were not attributed to differences in the configurations. For example, data distribution (or shards) must be the same in both solutions, as shards can greatly impact performance.

---

<sup>2</sup> A confidence level (CI) represents a range of estimates for a parameter of interest, computed at a designated CI (such as 90%). This statistic represents the long-run proportion of CIs (at the given CI) that theoretically contains the true value of the parameter.

<sup>3</sup> A one-tailed t-test with p-value = 0.1 was used to determine whether differences in performance were statistically significant.

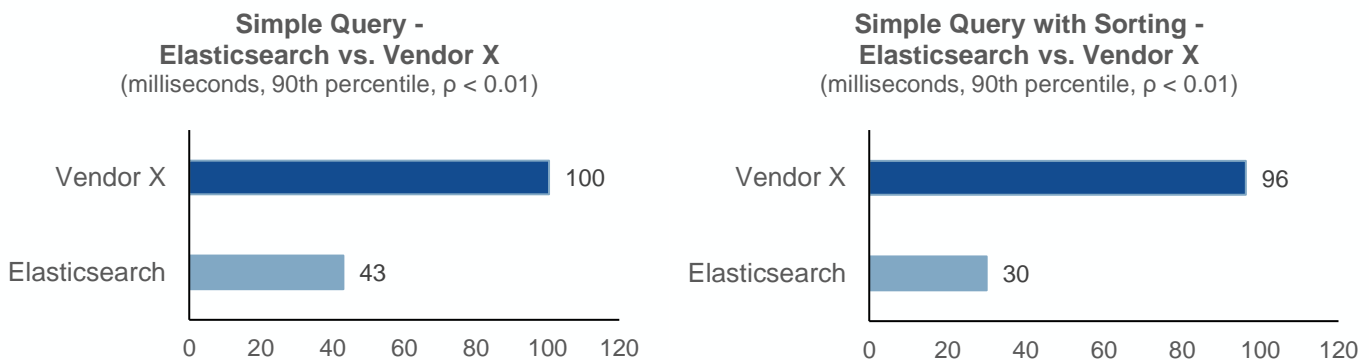
## Simple Query and Sort

To begin our evaluation, Enterprise Strategy Group first reviewed estimated times to complete simple text string queries, then sorted those results.

### Enterprise Strategy Group Testing

Test runs for each query represented 26,600 and 13,300 requests, respectively. Comparison of 90th percentile values for both query types, shown in Figure 3, revealed that Elasticsearch outperformed Vendor X.

**Figure 3.** Simple Query and Sort – Elasticsearch versus Vendor X



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

When considering the normalized data associated with both queries, Enterprise Strategy Group observed that Elasticsearch was estimated to be 76% and 140% faster than Vendor X, respectively.

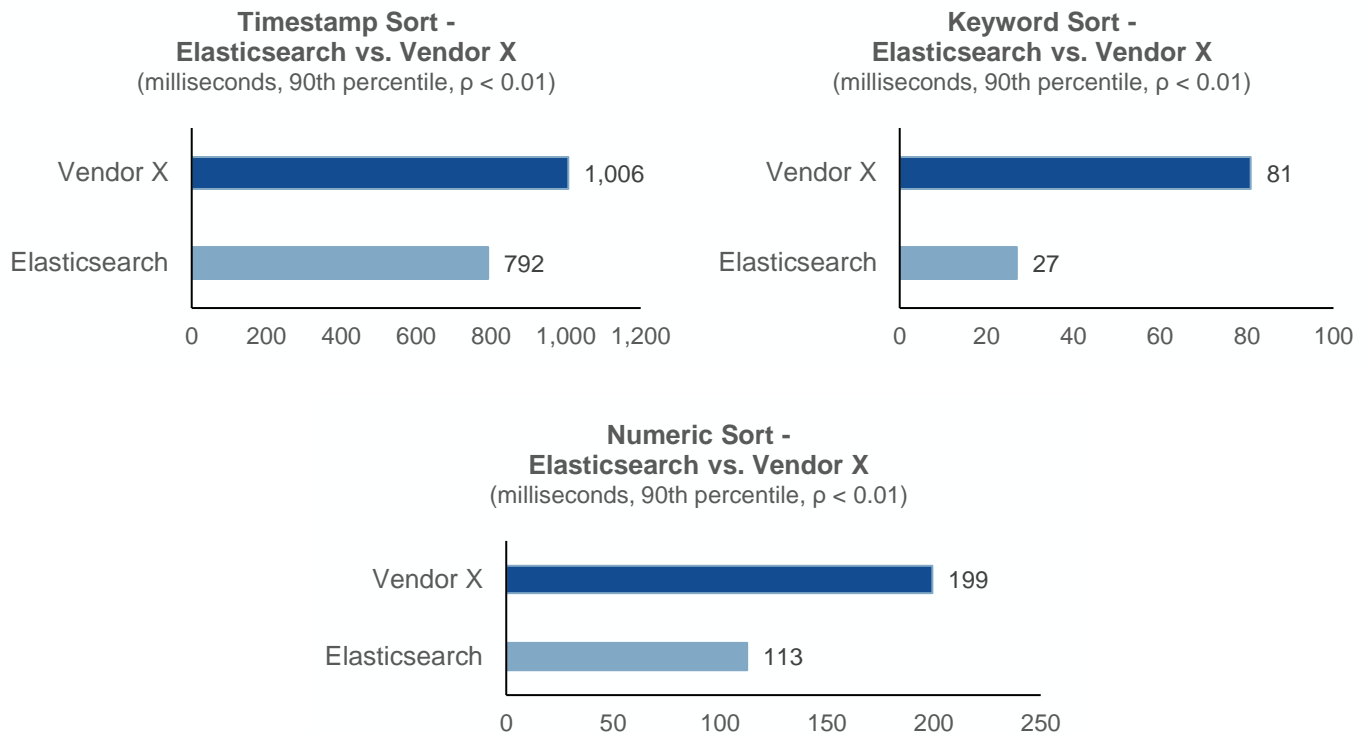
## Sorting with Timestamp, Keyword, and Numeric Data

Sorting data—alphabetically, numerically, chronologically—is used for supporting multiple business applications. End users of an e-commerce website sort search results (e.g., by manufacturer, price, overall review score, etc.) to easily discern their options. By quickly supplying relevant results in the preferred order, organizations running e-commerce sites can potentially convert these search results to sales. For security and observability use cases, sorting data, generated by multiple systems, is crucial to efficiently identifying potential issues, threats, or correlations that indicate trends or generate insights, thus reducing risk.

### Enterprise Strategy Group Testing

Enterprise Strategy Group first reviewed results showing times to complete data sorts according to timestamp associated with the test run, keyword (country code), and numeric data (population in increasing order). Test runs represented 88,800 timestamp sorting requests, 25,200 keyword sort requests, and 50,000 numeric sort requests.

Comparison of 90th percentile values for each set of results, shown in Figure 4, revealed that Elasticsearch outperformed Vendor X.

**Figure 4.** Timestamp, Keyword, and Numeric Sorting – Elasticsearch versus Vendor X

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

When normalizing the results for each query type, Enterprise Strategy Group saw that at the 90th percentile:

- Sorting by timestamp on Elasticsearch was estimated to be 24% faster than Vendor X.
- Sorting by keyword on Elasticsearch was estimated to be 97% faster than Vendor X.
- Sorting by numeric data (in increasing order) was estimated to be 53% faster than Vendor X.

### Date Histogram Aggregation Using Timestamped Data

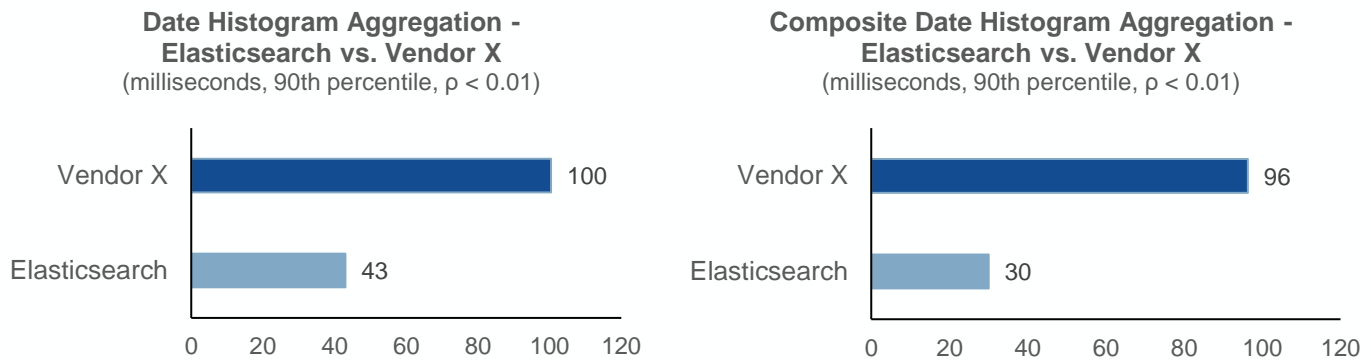
With a histogram based on time intervals, organizations can analyze website traffic by examining the frequency of occurrences during hours within the day, such as the number of visitors, content accessed, and responses to digital marketing campaigns (e.g., when prospects click on a call-to-action button). This histogram type can also categorize logs generated by IT assets, such as applications, servers, security devices, to figure out when peak usage or potential security attacks occur. However, generating histograms with usable time intervals, then correctly categorizing data according to those intervals, requires high search performance so that results are returned quickly.

### Enterprise Strategy Group Testing

Enterprise Strategy Group reviewed results showing times to complete date histograms, grouping data by hour, then minute. We also reviewed times for generating histograms that categorize data by hour and minute according to the day. Test runs represented 24,800 date histogram aggregation requests and 4,500 composite date histogram aggregation requests, respectively.

Comparison of 90th percentile values for both query types, shown in Figure 5, revealed that Elasticsearch outperformed Vendor X.

**Figure 5.** Date and Composite Date Histogram Aggregation – Elasticsearch versus Vendor X



*Source: Enterprise Strategy Group, a division of TechTarget, Inc.*

When considering the normalized data associated with both histogram aggregations, Enterprise Strategy Group observed that Elasticsearch was estimated to be 81% and 110% faster than Vendor X, respectively.

We noted that a composite data histogram is particularly useful when analyzing large sets of timestamped data. This bar chart “paginates” results by breaking down larger time intervals into smaller ones so that exploring data is completed more efficiently. For example, a security analyst can generate a histogram that counts event logs occurring during the days of a work week. To refine that analysis, histograms can then display results by day. With this level of granularity, the analyst can more easily pinpoint and track specific event occurrences when needed.

### Terms Query and Aggregation

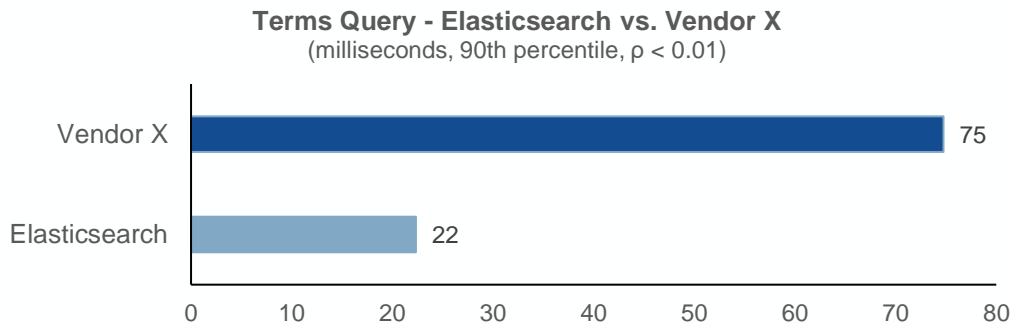
A terms query typically involves searching for specific words or terms amongst documents or web pages. End users can also request to generate a histogram categorizing data based on the aggregation of specific terms within an indexed data set. In both cases, fast response times are required when businesses conduct these queries to fulfill business objectives. Fast response time to terms queries and aggregations are required by customer support when they must find relevant documents that contain specific terms (such as product name and issue type) to provide relevant customer support or risk customer dissatisfaction. To uncover malicious activity, a security analyst may search for a specific IP address within log files, then search for the IT devices that recorded activity from the IP address.

When dealing with larger data sets, hundreds or thousands of unique terms may need to be analyzed before returning bucketized results based on individual terms or composite aggregations. Given the amount of statistical analysis to be done for analyzing term frequency and distribution, returning results quickly is critical.

### Enterprise Strategy Group Testing

Enterprise Strategy Group reviewed query times to conduct a search on a specific word or term within unstructured text. Test runs represented 13,700 requests. At the 90th percentile, the results showed that Elasticsearch outperformed Vendor X (see Figure 6). When considering the normalized data, Elasticsearch completed terms queries 108% faster.

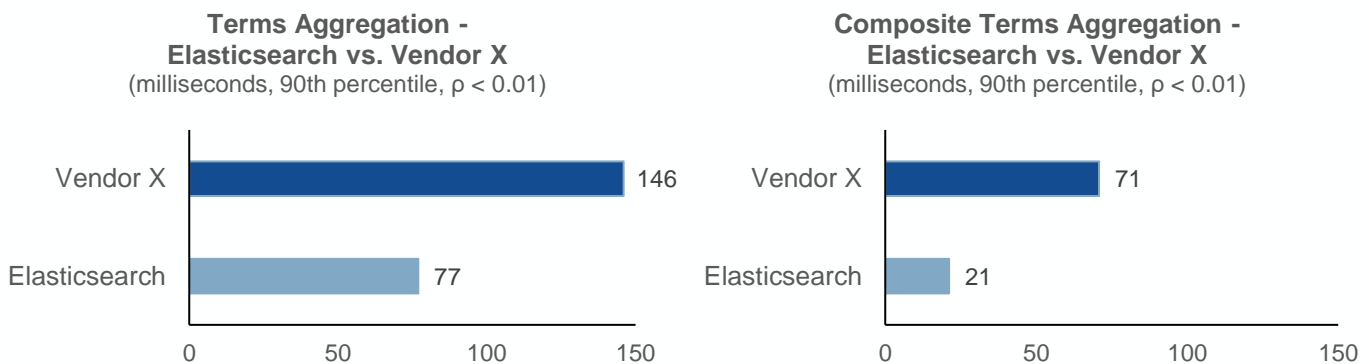


**Figure 6.** Terms Query – Elasticsearch versus Vendor X

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

We then conducted a total of 122 runs for 10 different requests on each search platform. These test runs represented 21,200 terms aggregation requests and 4,500 composite terms aggregation requests.

Comparison of 90th percentile values, shown in Figure 7, revealed that Elasticsearch's completion times were shorter than those of Vendor X. When considering normalized data, Enterprise Strategy Group estimated Elasticsearch to be 61% faster for term aggregations and 103% faster for composite terms aggregations, respectively.

**Figure 7.** Terms Aggregation, Individual and Composite – Elasticsearch versus Vendor X

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

## Ranges

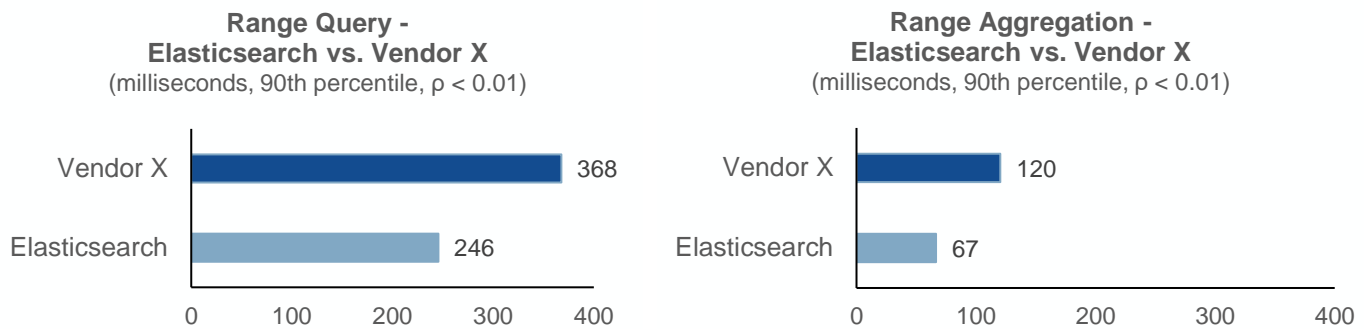
A range query extracts data from a select group of search results based on a specific range of values of an associated field. When searching for a house in a given city, an end user can specify a price range to narrow down the selected listings. A security analyst can also use range queries to narrow down a specific set of data based on time ranges (e.g., showing events occurring between 9AM and 10AM PT) to locate a specific anomaly.

## Enterprise Strategy Group Testing

Enterprise Strategy Group reviewed test results estimating completion times for complete range queries and aggregations. A total of 101,900 range requests and 12,400 range aggregation requests were completed.

Comparison of 90th percentile values, shown in Figure 8, revealed that Elasticsearch's completion times were shorter than those of Vendor X. When considering normalized data, Enterprise Strategy Group estimated Elasticsearch to be 42% faster than Vendor X for range queries. For range aggregations, Elasticsearch was estimated to be 66% faster.

**Figure 8.** Range Query and Aggregation Performance – Elasticsearch versus Vendor X



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

## Why This Matters

From generating revenue via an e-commerce website, to helping customer support resolve issues in real time, to supporting security analysts in identifying potential threats and anomalies, organizations rely on search applications to search for and deliver relevant data as quickly as possible, regardless of where that data lies or how large the data sets grow. Any delay in returning relevant results can lead to lost opportunities to increase revenue and decrease costs and risk. Such applications require a highly performant search platform.

Enterprise Strategy Group validated that Elasticsearch can query data, based on different parameters, more quickly than a comparable search engine offered by Vendor X. We reviewed times for completing a variety of queries and aggregations that typically characterize search performance— simple text search and sorting, sorting data in increasing or alphabetical order, generating regular and composite histograms, searching for and categorizing data based on unique terms, and aggregating data based on value ranges. Across all tests, we found that completion times were consistently faster than those produced by Vendor X. To account for outliers during test runs, we evaluated completion times using the 90th percentile. Differences in completion times, across all query types, were also statistically significant.

## Resource Utilization

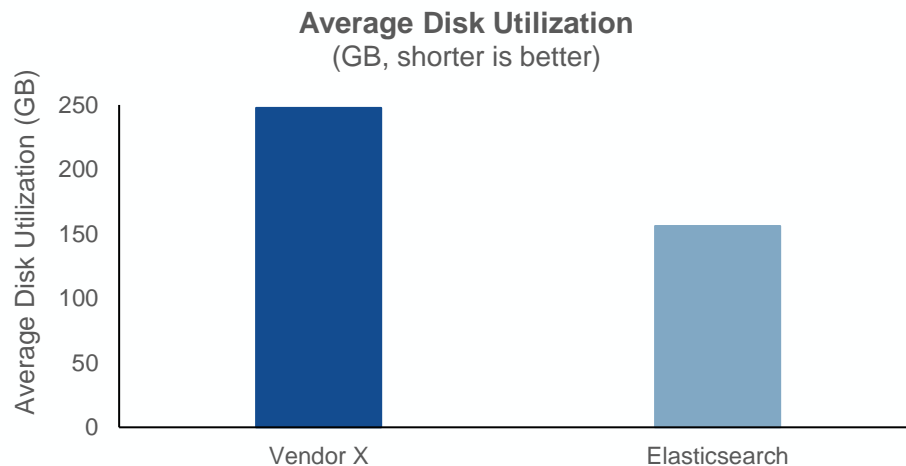
Achieving optimal performance of any IT asset should be done at the lowest cost possible. This is no different when evaluating search engine platforms. Elasticsearch has been designed to minimize compute resource utilization without sacrificing performance.

## Enterprise Strategy Group Testing

Finally, Enterprise Strategy Group reviewed results that estimated the average disk utilization across each cluster running Elasticsearch and the alternative offering. Each cluster ingested up to 1TB of log data. Total disk storage within each cluster was 1.5 TB. Out-of-the-box installations were performed with no optimizations or additional configurations.

Disk utilization was measured across each cluster as both search platforms ingested more data for each test run until ingesting the entire 1TB (e.g., beginning at 100GB and doubling the amount for subsequent runs). When comparing the average disk utilization across all test runs, Enterprise Strategy Group estimated that Elasticsearch used 37% less disk storage than Vendor X did (see Figure 9).

**Figure 9.** Average Disk Utilization – Elasticsearch versus Vendor X



Source: Enterprise Strategy Group, a division of TechTarget, Inc..

### Why This Matters

Achieving optimal performance of search applications is a must, especially when fulfilling business needs. Yet, with the ongoing strain in budget dollars, organizations must remain careful when investing limited funds. Maintaining high performance/cost ratios remains a concern.

Enterprise Strategy Group validated that Elasticsearch can consume less IT resources when running on Google Cloud-based Kubernetes clusters. We observed that, on average, Elasticsearch can ingest up to 1TB of log data with 37% less disk storage than the alternative offering. Less disk storage translates directly into less cloud resources expenses.

## Conclusion

Enterprise Strategy Group research found that some of the considerations that respondents reported that they will use to justify technology investments in 2023 include improving their customer experience (32%), business process (28%), and digital collaboration capabilities (28%).<sup>4</sup> One approach to achieving these improvements is to use a highly performant search platform that supports end users in organizations to obtain critical data quickly or that delivers relevant data to external customers. Maximizing search performance—delivering relevant results within seconds—can optimize the user experience. Good user experiences of search applications can ultimately help organizations achieve multiple business objectives, including increasing revenue, maximizing employee productivity, decreasing costs, and minimizing risk.

<sup>4</sup> Source: Enterprise Strategy Group Research Report, [2023 Technology Spending Intentions Survey](#), November 2022.

To maximize the performance of search applications, organizations can leverage Elasticsearch, a distributed analytics engine designed to work with all types of data obtained from multiple and disparate data stores on premises and in the public cloud. Elasticsearch has designed the platform to search for, discover, analyze, and index the most relevant results for any query, regardless of volume of data ingested. Because Elasticsearch can run either on premises or on any public cloud, organizations have the flexibility to deploy the platform as business needs dictate. Organizations that partner with Elastic also have access to an established community of contributors, experts, and customer support professionals to ensure that applications are leveraging the latest technology development and resolving service-affecting issues as quickly as possible. With Elasticsearch, organizations can ensure that data queries deliver relevant results quickly in search, security, and observability use cases.

Enterprise Strategy Group validated that Elasticsearch delivers better performance and consumes less resources than an alternative search platform. We reviewed the performance of conducting various query types and average disk utilization rates on both search platforms. Based on statistical methods and analysis, we estimated that Elasticsearch was:

- 76% faster in simple text queries.
- 140% faster in sorting results of simple text queries.
- 24%, 97%, and 53% faster in timestamp, keyword, and numeric sort queries, respectively.
- 81% and 110% faster in date histogram and composite date histogram aggregations , respectively.
- 108% faster in terms queries.
- 61% and 103% faster in terms and composite terms aggregations, respectively.
- 42% and 66% faster for range queries and range aggregations, respectively.

Enterprise Strategy Group also validated that Elasticsearch consumes an average of 37% less disk storage of a five-node GCP Kubernetes cluster than the alternative offering.

End users, both internal and external to an organization, know the value of high-performance search applications. Anything less than optimal performance can quickly lead to undesirable business results. Should you be looking for a partner that can provide a search platform to optimize internal and external-facing search applications, we suggest closely looking at Elasticsearch.

©TechTarget, Inc. or its subsidiaries. All rights reserved. TechTarget, and the TechTarget logo, are trademarks or registered trademarks of TechTarget, Inc. and are registered in jurisdictions worldwide. Other product and service names and logos, including for BrightTALK, Xtelligent, and the Enterprise Strategy Group might be trademarks of TechTarget or its subsidiaries. All other trademarks, logos and brand names are the property of their respective owners.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.


Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at [cr@esg-global.com](mailto:cr@esg-global.com).

---

**About Enterprise Strategy Group**

TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

 [contact@esg-global.com](mailto:contact@esg-global.com)

 [www.esg-global.com](http://www.esg-global.com)