aws marketplace

Vespa

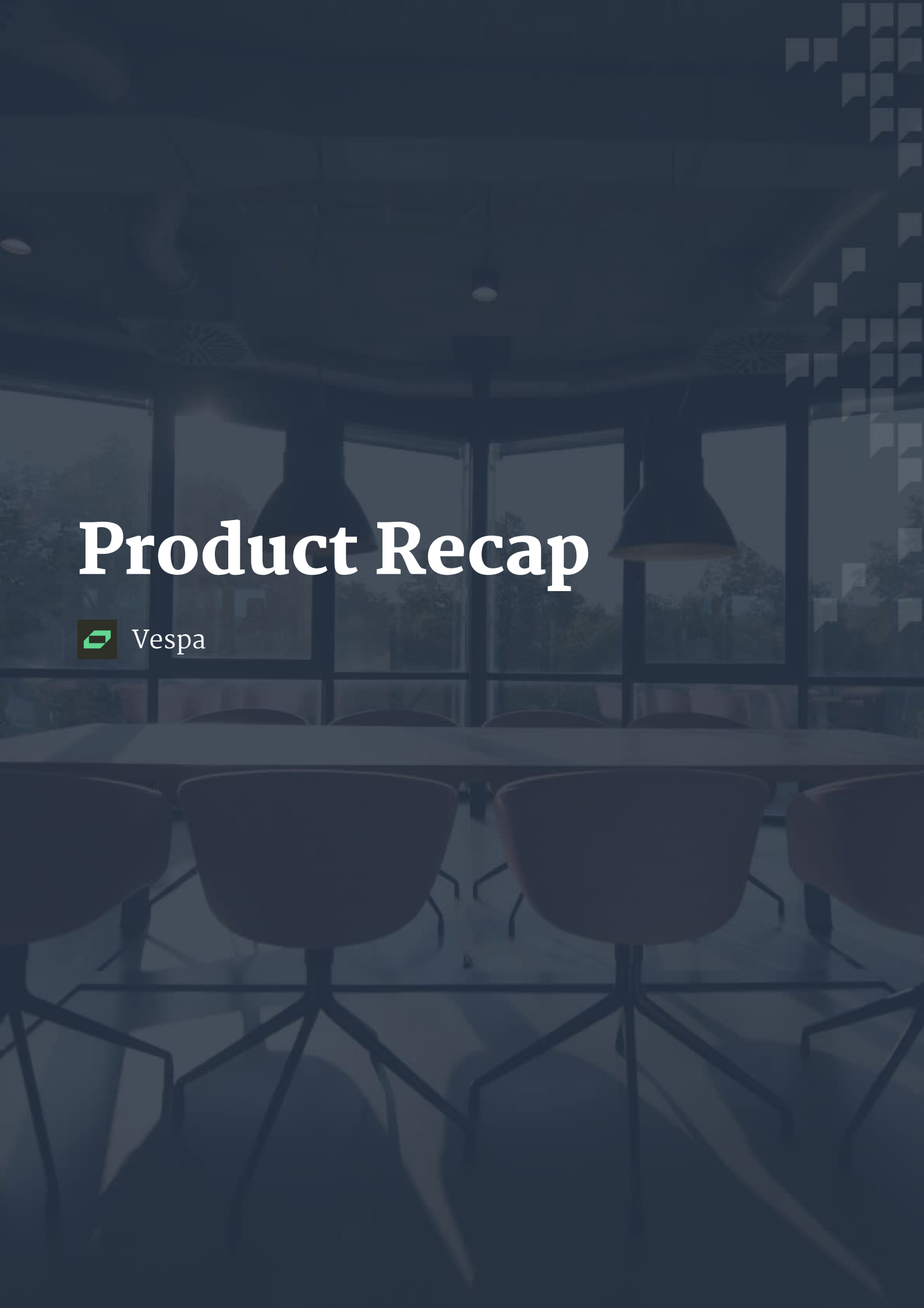
Reviews, tips, and advice from real users






Powered by PeerSpot

Contents



Product Recap

Vespa

Vespa Recap

Vespa is a versatile product that enhances search functionality and improves the performance of large-scale applications. Users have reported using Vespa for content recommendation, personalization, and real-time analytics.

It handles high volumes of data and delivers fast and accurate search results. Vespa is also valuable for building intelligent applications, powering e-commerce platforms, and enabling efficient data retrieval and processing.

With exceptional performance, efficient fuel consumption, stylish design, comfortable seating, and smooth handling, Vespa is the complete package.

Valuable Features

Excerpts from real customer reviews on PeerSpot:

- "While conducting A/B testing, Vespa seemed to be performing slightly better than Elasticsearch, especially in search relevancy within live production systems, and its performance was decent."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

- "The best feature to me is the LTR feature, the ranking feature to be specific."

Kelyn Ukiru
Recommendation Platform Developer at Remote Labs

- "Vespa is very good and it improves our product, and we got more clients."

Shubhang Dutta
Integration Related To Ai at RedBlink Technologies

"The most outstanding features and characteristics of Vespa include an architecture that lets you focus on implementing features, the function that automatically manages sharding and shards is excellent, and the flexibility of the server cluster and infrastructure architecture is outstanding."

Verified user

검색 엔지니어 at a computer software company with 1,001-5,000 employees

What users had to say about valuable features:

"The most outstanding features and characteristics of Vespa include an architecture that lets you focus on implementing features. The function that automatically manages sharding and shards is excellent, and the flexibility of the server cluster and infrastructure architecture is outstanding. The indexing performance was greatly improved by using HTTP/2.

"In the pilot phase, I expected that, rather than organizational changes, in terms of service on the actual business platform and from the perspective of customer usability, the UI features and planning features could have been provided differently if it had been adopted.."

Verified user

검색 엔지니어 at a computer software company with 1,001-5,000 employees

Read full review

"One of the best features that Vespa offers is natively handling embeddings or vectors, along with its capability for really fast searches. The powerful DSL provided by Vespa allows you to define search calculations, which I used extensively over the period of one year.

"Vector search is definitely the biggest selling point for Vespa. Even though Elasticsearch has vector search capabilities, it is not as powerful as what Vespa offers. Since Vespa natively supports sparse embeddings, that was an advantage, but in the end, I opted for dense embeddings using Google's Gemma's embeddings, as I worked on retraining that model on our dataset.

"Vespa's DSL includes many built-in functions, which is quite powerful. Even without embedding features, the DSL shines. However, it takes considerable time for indexing—initial runs took almost a whole day to create the embeddings and push them into Vespa, and it required adequate resources to run. Another issue was its inability to handle synonyms in the same manner as Elasticsearch.."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

Read full review

"The best feature to me is the LTR feature, the ranking feature to be specific. For most other vector databases, you perform the query and then apply your own logic on LTR outside the database, but with Vespa, these two operations can be done within the database. This means that latency is reduced and bottlenecks are reduced if these two operations can be done within the database. It is more having your filter and sort operation within the database itself. I also appreciate the filter part that Vespa offers.

"The results have been better with Vespa. The matching has been far better with Vespa. Before, I was using other solutions such as PGVector and Pinecone and Weaviate, but they were harder to integrate, meaning more work for the developer. With Vespa, the integration on the development side has been easier.

"I cannot say that the improved results are directly tied to Vespa alone because many iterations have been made, and this includes the general architecture. The fact that Vespa does its own indexing, and I can just receive a string of text, index it and store it as a vector, is remarkable.."

Kelyn Ukiru
Recommendation Platform Developer at Remote Labs

Read full review

"Earlier we used Quadrant where there is only vector embeddings and search on that basis. Vespa provided us a highly scalable and more reliable platform. Vespa also provides BM25 text search and embedding search. The main reason to move to Vespa is hybrid search.

"We have explored BM25 and hybrid search. We have implemented direct search and also embedding search by creating embeddings and storing them in Vespa. We can also use direct text search based on the user's query. This way we have implemented hybrid search and get the user's response.

"It works very well because we have many documents, plus a single document is also very long. Vespa is very good with retrieval and high-volume queries.

"Earlier we used Quadrant and now we are moving with Vespa. In Quadrant, we have a concept of collection where for every assistant, we create a new collection, and every document in that assistant goes to that particular collection. In Vespa, there is no concept of collection so we have to separate it on the basis of that assistant. That makes it unique. We were familiar with only the single, single, single collection for that specific assistant. With Vespa, we have all in one place and get it separated out on the basis of assistants and the environment we are using.."

Shubhang Dutta
Integration Related To Ai at RedBlink Technologies

Read full review

Other Solutions Considered

"I was using PGVector before switching to Vespa. The reason that made me switch to Vespa is because I needed more functionality, such as the ranking feature. There were some other options on the table; Weaviate was one of them.."

Kelyn Ukiru
Recommendation Platform Developer at Remote Labs

Read full review

"Before adopting Vespa, we were considering other solutions like Qdrant and Milvus, and we decided to evaluate Vespa because we felt those two solutions were somewhat lacking in terms of scalability.."

Verified user
검색 엔지니어 at a computer software company with 1,001-5,000 employees

Read full review

"Before choosing Vespa, I explored a few other search engine solutions, starting with Orama, a Node.js and TypeScript-based search engine that struggled to handle six million SKUs, and then Typesense, which aimed for instant searches but failed to accommodate the numerous attributes I needed for sparse data. That led me to Vespa, which met my expectations.."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

Read full review

ROI

Real user quotes about their ROI:

"I would not agree with seeing a return on investment since we had not fully deployed Vespa into production, with only two people working on it for approximately a year and a half, which did not require a large team. However, we spent about two thousand to three thousand dollars per month on AWS while using Vespa, which was higher compared to around one thousand to one thousand five hundred dollars per month for Elasticsearch, although we saw some slight improvements in key metrics before stopping the A/B test.."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

Read full review

Use Case

"I use Vespa as a vector database for ranking and matching. I have jobs and candidates indexed in Vespa, which is a vector database. When I have a job and need to get the first 50 candidates who match that job, a normal vector search would not retrieve the first 50 because I may need to filter or rank based on some features and fields. Vespa helps by allowing me to first select the first 200, and then within the 200, I rank the first 50 based on certain criteria.."

Kelyn Ukiru
Recommendation Platform Developer at Remote Labs

Read full review

"My main use case for Vespa is implementing it as the back-end search engine for an e-commerce site, where we have about six million products, or six million SKUs, that we are selling. I implemented Vespa as an alternative for Elasticsearch.

"Using Vespa for the e-commerce site involved utilizing it as the backend search engine to replace Elasticsearch, which we felt was not doing us justice. The very first thing I did was convince my CIO to try out Vespa. We did a quick proof of concept, engaged with the right people through the Vespa Slack channels, and then we did the actual implementation, including A/B testing it against the previously running fully optimized Elasticsearch pipeline.."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

Read full review

“I have been using Vespa at a POC level, and we ran it for a period of roughly less than a month.

“The main purpose of using Vespa was to run a POC to introduce semantic search into the company, and we reviewed and researched various solutions available on the market. As a result, Vespa seemed interesting. The company was using Elasticsearch, and we were interested in how it compared to Elasticsearch in terms of infrastructure architecture, usage purpose, and handling embeddings within the engine in the way we wanted, so we decided to try Vespa.

“For the query part, we focused on creating queries using embeddings with the E5 Multilingual Embedder on top of the existing lexical search of product titles and product details for semantic search. We wrote the queries with the purpose of providing hybrid search.

“My company deployed Vespa in a cloud environment..”

Verified user

검색 엔지니어 at a computer software company with 1,001-5,000 employees

Read full review

“My full name is Shubhank and I am serving in Redblink Technologies in Mohali where I have been doing integration work related to AI. I implemented RAG.

“The previous year, we were using Quadrant as a vector store. With that, we were creating many collections there. Our company discussed internally and decided to move to Vespa. This was about six or eight months ago. We are using Vespa in our RAG pipeline.

“We have implemented a RAG pipeline where we have document retrieval. Users can chat with their documents. We are breaking down our documents into meaningful chunks using LangChain4j and feeding that directly into Vespa as a vector store. Later, while the user chats or starts a chat with the document, we can retrieve according to the user's prompt.

“We have such more use cases. We have a client, CPA Pilot, where there are many text documents, so we directly chunk those documents. There are very large documents, so in Quadrant, the collections were almost full. Inside Vespa, there is no system of collections, so that also helped us. We use self-hosted Vespa for that particular client and we are chunking down the long documents using LangChain4j and hitting Vespa to store it. During retrieval, we get good results and get proper relevant scores based on the user's query..”

Shubhang Dutta
Integration Related To Ai at RedBlink Technologies

Read full review

Setup

The setup process involves configuring and preparing the product or service for use, which may include tasks such as installation, account creation, initial configuration, and troubleshooting any issues that may arise. Below you can find real user quotes about the setup process.

"The setup cost is definitely huge, and pricing is also steep. In terms of licensing, it seems generous for those who do not want to engage with Vespa's hosted services.."

Ganaraj Amakrishna

Lead Technical Architect at Zoro UK

Read full review

"AWS provides us with more analytics with Vespa, such as how it is performing on the servers. It is easy because of their well-documented documentation.

"We are using self-hosted Vespa in our AWS servers.

"The setup process is fine. It helps us save money and we got very good responses from the users.."

Shubhang Dutta

Integration Related To Ai at RedBlink Technologies

Read full review

Customer Service and Support

"The customer support I received was pretty good, mainly through interactions in the Slack community, where I typically got responses within hours or by the next day, leading me to rate them an eight or maybe even nine.."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

Read full review

"I do not have that much involvement with the customer support. I have raised some questions on Slack for the Vespa community and received responses in 24 hours. I have discussed my concerns and questions. The community support is very good.."

Shubhang Dutta
Integration Related To Ai at RedBlink Technologies

Read full review

Other Advice

“Vespa had a positive impact on our company, but we do not have any cases of actual production application.

“I hope that Vespa will surpass Elasticsearch in the global AI embedder and embedding market, and in the AI DB and vector DB market. I gave this review a rating of eight out of ten..”

Verified user
검색 엔지니어 at a computer software company with 1,001-5,000 employees

Read full review

“Up to now, I am still in the building phase. I have not gone commercial with my product, and so I cannot give a relevant answer about that. I am still trying out Vespa to see if it actually meets my business need. I would tell others that the product is actually good if they have some resources on their side because it is resource-intensive. It actually requires someone who knows what they are doing to reap most of the benefits out of Vespa because you do not have to implement most of the features in the code layer; you can just do it at the database layer. I would rate this product an 8 out of 10..”

Kelyn Ukiru
Recommendation Platform Developer at Remote Labs

Read full review

“For anyone who wants to use a vector store, they should do research on their end, and if nothing comes up after discussion and research, I recommend using Vespa

because they have good reliability. The main thing is the speed. The retrieval speed is very good. I recommend Vespa for systems to get integrated with.

“Vespa is very good and it improves our product, and we got more clients. We got very good results and very good relevance. This mainly depends on how you can design the Vespa document schemas. The document schema design determines how your relevance will come and how your retrieval will be done. The feedback for how Vespa responds is good and also fast. We are using Amazon Web Services (AWS) and it is easy because of their well-documented documentation. I give Vespa a rating of nine out of ten..”

Shubhang Dutta
Integration Related To Ai at RedBlink Technologies

Read full review

"I would rate Vespa a six, as it is a powerful tool with great potential in terms of search engine capabilities, but the steep learning curve and initial setup costs are significant downsides.

"I chose six because of the steep learning curve and the substantial initial costs involved with setting up Vespa. If it were feasible for people with limited budgets, even as low as fifty dollars a month, it would be more appealing.

"While conducting A/B testing, Vespa seemed to be performing slightly better than Elasticsearch, especially in search relevancy within live production systems, and its performance was decent. Comparing raw Elasticsearch text-based search against Vespa's vector-based and text-based search, we were already recommending Vespa to several peer companies.

"During A/B testing, looking at conversion rates, search-to-basket ratios, and add-to-basket ratios showed improvement until we shut it down. It took several iterations to get the results, particularly after switching to Embedding Gemma, emphasizing that the quality of embedding used heavily influenced the outcome.

"Nothing else comes to mind regarding improvements needed for Vespa.

"I would not suggest Vespa unless you are an enterprise due to the steep learning curve and significant infrastructure costs involved. My overall rating for Vespa is six out of ten.."

Ganaraj Amakrishna
Lead Technical Architect at Zoro UK

Read full review

Top Industries

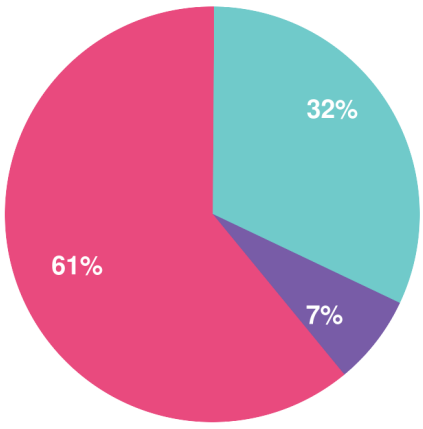
by visitors reading reviews

Computer Software Company	16%
Comms Service Provider	12%
Financial Services Firm	10%
Healthcare Company	8%

Company Size

by reviewers

by visitors reading reviews



Large Enterprise Midsize Enterprise Small Business

About this buyer's guide

Thanks for downloading this PeerSpot report.

The summaries, overviews and recaps in this report are all based on real user feedback and reviews collected by PeerSpot's team. Every reviewer on PeerSpot has been authenticated with our triple authentication process. This is done to ensure that every review provided is an unbiased review from a real user.

Get a custom version of this report... Personalized for you!

Please note that this is a generic report based on reviews and opinions from the collective PeerSpot community. We offer a customized report of solutions recommended for you based on:

- Your industry
- Company size
- Which solutions you're already considering

The customized report will include recommendations for you based on what other people like you are using and researching.

Answer a few questions in our short wizard to get your customized report.
Get your personalized report here

About PeerSpot

PeerSpot is the leading review site for cloud, AI, and business software. We created PeerSpot to provide a trusted platform to share information about software, applications, and services. Since 2012, over 22 million people have used PeerSpot to choose the right software for their business.

PeerSpot helps tech professionals by providing:

- A list of products recommended by real users
- In-depth reviews, including pros and cons
- Specific information to help you choose the best vendor for your needs

Use PeerSpot to:

- Read and post reviews of products
- Access over 30,000 buyer's guides and comparison reports
- Request or share information about functionality, quality, and pricing

Join PeerSpot to connect with peers to help you:

- Get immediate answers to questions
- Validate vendor claims
- Exchange tips for getting the best deals with vendor

Visit PeerSpot: www.peerspot.com

PeerSpot

244 5th Avenue, Suite R-230 • New York, NY 10001

reports@peerspot.com

+1 646.328.1944