

Taboola

Vertica delivers a fully scalable platform for sophisticated data analytics with higher user concurrency and performance.

Overview

Taboola is the leading content discovery platform, serving more than 450 billion recommendations of articles, blogs, videos, products, and apps to more than one and a half billion unique users every month on thousands of premium sites and mobile carriers. Publishers, brand marketers, and performance advertisers leverage Taboola to retain users on their sites, monetize their traffic, and distribute organic and sponsored content as well as video to engage high-quality audiences.

Challenge

For Taboola, the name of the game is growth. In a recent year, page views increased by 36 percent, the number of recommendations served

to the audience by 34 percent, and the incoming data volume by 220%. Taboola works with media firms and content publishers to build an engaged audience and increase revenue for its partners, such as CBS Interactive, Euronews, Whirlpool, and InnoGames. When site visitors click on articles, their browsing history is analyzed to give them personalized recommendations that extend their browsing time on the site and the provider's engagement with them. Keren Bartal, Director of Data Engineering for Taboola, explains how the explosive growth of the Taboola platform quickly became a challenge: "When Taboola first launched, we used a traditional relational database. But, with our growth rate at orders of magnitude, and our users generating upwards of 30 TB data each day, our database strategy was limiting our ability to provide value to our partners."

To improve its analytical performance, Taboola decided to move to a columnar database. Bartal and the team were looking for faster query time, flexibility to add to and extend the data, a scalable platform that would keep up with Taboola's fast predicted growth, a predictable cost model, and the ability to cope with unexpected data center or server downtime without service interruption. The team was fully aware that this was a tall order, but it was the only way Taboola could honor its own service commitments to partners.

"With Vertica as the compute engine at the heart of our data, we can now offer a fully scalable analytics offering with higher user concurrency and performance. Our partnership is set to take us to exciting places."

KEREN BARTAL

Director of Data Engineering
Taboola



Taboola

At a Glance

- **Industry**
Software and Technology
- **Location**
Tel Aviv, Israel
- **Challenge**
Produce relevant, accurate, and timely reports by analyzing more than 30 TB of new data each day.
- **Products and Services**
Vertica Analytics Platform
- **Results**
 - + Faster data loading and processing
 - + Scalable solution to follow Taboola's predicted growth
 - + Increased query speed performance
 - + Effective partnership to develop new Vertica features

“Our analysts love the fact that Vertica is integrated with a number of data visualization tools of their choice, such as Tableau and QlikView.”

KEREN BARTAL

Director of Data Engineering
Taboola

Contact us at:
www.vertica.com

Like what you read? Share it.



Solution

After a thorough evaluation, Taboola chose the Vertica Analytics Platform. Bartal notes: “A proof-of-concept showed us a vast increase in query speed, which was decisive for us. As a company, our strategy is very much on-premise, as a cloud model is cost-prohibitive with our level of data volumes. Vertica seemed to fit our environment and culture very well and we could see ourselves up and running quickly. Vertica’s MPP (Massively Parallel Processing) architecture meant we could deploy a cluster of servers for much faster data loading and processing.”

The Vertica team supported Taboola through their architecture definition and implementation. Thanks to an effective team effort, Taboola achieved great results quickly. Vertica was deployed on multiple backend and frontend clusters to isolate the different workloads. The backend clusters are used for continuous, heavy aggregations of raw data and reports; each day, Vertica ingests up to 500GB of compressed raw data. More than 5,000 daily reports are generated from this. The backend clusters also integrate third-party data with Taboola data for more sophisticated Business Intelligence (BI) analysis.

To enrich the data, workloads from Google BigQuery are also aggregated and loaded into Vertica. This integration capability is appreciated, as Bartal comments: “We like that Vertica can slot into our existing infrastructure. We have a rich ecosystem with open source technologies such as Apache Hadoop, Kafka, and Spark, all of which are supported by Vertica.” Bartal also noted a preference for Vertica by the Taboola analyst community: “Our analysts

love that Vertica is integrated with a number of data visualization tools of their choice, such as Tableau and QlikView.”

Reports for Taboola partners need to be relevant, accurate, and timely, and the company always looks to exceed partners’ expectations and deduce more from the data than they might have thought of themselves. The data needs to be as “fresh” as possible when it’s analyzed, and so the faster the data is loaded and computed, the faster partners receive the intelligence to help them fine-tune their content offer or their marketing campaigns and deliver a more compelling user experience.

A Vertica cluster powers a dashboard for Taboola partners. This is accommodated in UI and API. Offering both is particularly important as the UI is used by analysts and business partners to generate pre-formatted and custom dashboards, with user-friendly drag-and-drop features for ad-hoc analysis. The dashboard system pulls directly from Vertica, which really improves the data quality and delivery speed. APIs are used when a partner’s IT systems interact directly and automatically with the Vertica data. The aggregated data is automatically retrieved so that a partner’s marketing campaigns can be dynamically adjusted as a result of continuous data analytics or integrated into billing and report platforms. Vertica connectors help Taboola integrate with Hadoop HDFS, which holds the raw data. Thousands of concurrent users access the various frontend clusters at any given time, which is why platform stability and scalability were such important factors in the decision to choose Vertica.

Multiple clusters have the potential to introduce many challenges around data consistency and replication. These were met with a lot of in-house tool development, as Bartal explains: “We developed our own ETL tool to manage integrations between Vertica and our other platforms, such as MySQL, Cassandra, various APIs, and cloud services.”

Results

Vertica offers the flexibility that Bartal sought: “Vertica shows me a vast variety of metrics; everything is completely visible and accessible. Through encoding, we continuously tune and improve our projections, which speeds up our query runtime as a result.”

One of the best experiences with Vertica has been the partnership that has developed over the years, according to Bartal: “The Vertica partnership is key to us. We have really pushed the boundaries of the solution and the Vertica team has been there to support us every step of the way. This is showcased particularly when we see our discussions result in new Vertica features. We are excited to be a Beta customer for Vertica in Eon Mode on-premise. This will give us the opportunity to take advantage of flexible compute and storage resources, similar to a cloud model, to manage dynamic workloads.”

She concludes: “Our phenomenal growth meant we needed robust and enterprise-ready support in data analytics. Our vision is to reach more people, via more channels. With Vertica as the compute engine at the heart of our data, we can now offer a fully scalable analytics offering with higher user concurrency and performance. Our partnership is set to take us to exciting places.”