#BLOG 1 | Implementing Vector Similarity Search in GCP's Database Ecosystem

OUR GCP EXPERTS HAVE A DEEP DIVE ON THE CURRENT SUPPORT FOR VECTOR SEARCH TO POWER AI-DRIVEN SEARCH APPLICATIONS

SWORD
UPGRADE YOUR BUSINESS

Table of Contents

1   Vector Similarity Search in PostgreSQL

Since June 2023, Google has integrated vector similarity capabilities into its PostgreSQL offerings, specifically within CloudSQL for PostgreSQL and AlloyDB for PostgreSQL. This initiative uses the pgvector extension, which Google updates frequently, aligning closely with the latest releases. For instance, pgvector v0.6.0, which introduced parallel HNSW (Hierarchical Navigable Small World) index builds in January 2024, was promptly updated in CloudSQL by April 2024.

Here is an example of how to leverage HNSW index for fast vector similarity search on CloudSQL for PostgreSQL.

Additionally, via the PostgreSQL extension google_ml_integration that Google maintains, you can grant CloudSQL and AlloyDB IAM permissions to call Vertex AI directly.

2   BigQuery and Vector Support

While vector support in BigQuery isn't generally available as of July 2024, its development is highly anticipated. BigQuery is primarily used as a data warehouse solution, and vector support will allow users to perform Approximate Nearest Neighbour (ANN) searches using a simple function call to VECTOR_SEARCH . This functionality will significantly enhance the analytical capabilities of applications by providing both approximate and exact search results within the vast datasets typically stored in BigQuery.

3   AlloyDB: A Fusion of Operational and Analytical Workloads

AlloyDB, which became generally available in March 2024, merges operational and analytical database capabilities within a PostgreSQL- compatible, fully managed service. This product is ideal for demanding transactional workloads with built-in AI features, positioning it as a strong competitor against services like AWS Aurora and MySQL HeatWave.

4 Unique Features of AlloyDB

One of the features of AlloyDB is the separation of computing and storage, using the same robust storage technology, Colossus, that powers BigQuery. This architectural design allows for scalable and flexible resource management.

Furthermore, AlloyDB enhances query performance through its columnar engine, which optimises analytical queries by storing data in a columnar format. These capabilities are supported by extensive RAM allocations, necessary for managing large volumes of data efficiently.

AlloyDB AI further extends these capabilities by integrating functionalities for building generative AI applications directly within the database using the pgvector extension. This feature supports SQL operations on vector embeddings and AI model inference, making AlloyDB a versatile tool for developers aiming to incorporate AI into their applications seamlessly.

While AlloyDB does not seek to replace BigQuery, which supports several-orders-of-magnitude larger analytical workloads, it is a step further that blends the storage layer between the two.

5 CloudSQL for PostgreSQL

CloudSQL for PostgreSQL is another integral part of Google's database ecosystem, specifically tailored for those seeking a fully managed relational database service that combines the rich feature-set of PostgreSQL with the flexibility and scalability of the cloud.

This service allows developers to leverage PostgreSQL's capabilities without the overhead of managing infrastructure, providing automatic backups, replication, and failover to ensure high availability and reliability. The recent addition of vector similarity search through the pgvector extension significantly enhances its utility by enabling advanced search functionalities directly within the database environment. This allows for more complex queries and data relationships to be managed effectively, facilitating sophisticated data interactions necessary for modern applications.

Additionally, Google's consistent updates to pgvector in CloudSQL, such as the introduction of parallel HNSW index builds, ensure that users have access to cutting-edge capabilities, making CloudSQL for PostgreSQL an increasingly attractive option for developers looking to build scalable, data-driven applications.

6 Conclusion

Google's strategic enhancements to its database offerings reflect a broader trend in the software industry towards more integrated and AI-enhanced data management solutions. By embedding vector similarity search capabilities into PostgreSQL and soon in BigQuery, Google is not only upping the stakes in database technology but also providing developers and businesses with powerful tools to harness the potential of their data fully. As these technologies become more mainstream, they promise to revolutionize how companies approach data analysis and decision-making processes.