Enterprise knowledge graphs (EKG) require graph databases, which serve multiple purposes. The engines must facilitate the advanced data integration and metadata data management scenarios where an EKG is used for data fabrics or otherwise serves as a data hub between diverse data and content management systems. The same engines are expected to efficiently deal with computationally challenging data analytics, discovering multi-hop relationships across networks of concepts, entities, assets, documents and other resources.

We are happy to announce that Ontotext GraphDB officially passed two benchmarks of the Linked Data Benchmarking Council (LDBC), namely the Social Network Benchmark (SNB) and the Semantic Publishing Benchmark (SPB). The audited results have been recently published on the corresponding webpages of LDBC, turning GraphDB into the only engine which is proven to deal efficiently with both graph analytics (SNB) and metadata management (SPB) workloads.

RDF engines are good for graph analytics

Historically, the Labeled Property Graph (LPG) engines were optimized to deal with graph analytics, while the Resource Description Framework (RDF) engines were designed for data publishing and metadata management. Since the beginning of LDBC, there was clear separation: RDF engines were audited only on SPB, while LPG (and other graph analytics-optimized designs) were audited only on SNB. This era is over! GraphDB officially passed SNB's Interactive Workload at scale factor 30 (SF30) – a graph of 1.5 billion edges.

The benchmark simulates analytical queries against social networks data – messages, comments, people related to other people, cities, universities, companies, etc. SNB is the most advanced graph analytics benchmark, the result of cooperation between the leading research groups in the field (e.g., CWI) and some of the major graph database vendors (e.g., neo4j). Its data generator creates a realistic graph, which is as diverse and challenging as possible.

The Interactive workload consists of 14 queries such as "People that a person is connected to at up to 3 steps via 'knowns' relationships" and "Find the shortest path between two persons". The workload also includes data updates and there are special provisions in the test to verify that the engine can deal with those in a consistent manner, according to the corresponding transaction isolation level.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

GraphDB was audited to perform 12 operations/second on an AWS r6id.8xlarge server (256GiB RAM, Intel Xeon Platinum 8375C) against a test driver configured with 4 read and 4 write threads. This achievement was possible due to GraphDB's Graph Path Search extension, introduced in 2021 and optimized several times since.

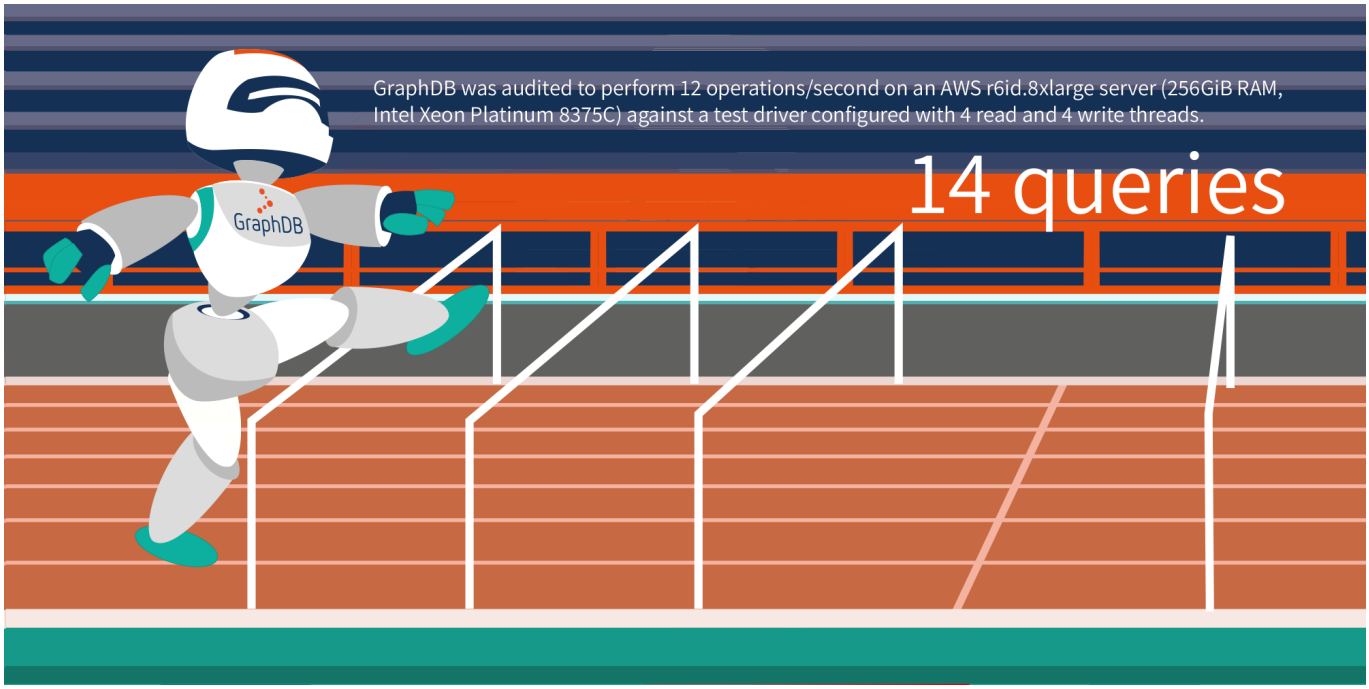
As expected, these SNB results do not match the performance of specialized graph analytics systems, such as TuGraph, that implements SNB via stored procedures written in C++, rather than via standard query language. Still, GraphDB's results are the only ones where a general purpose database engine passes the benchmark without custom indices or compression tailor made for this benchmark!

Scalable handling of concurrent clients using multiple CPU cores

It is every vendor's mission to offer to its users optimal performance across all relevant workloads. This usually boils down to two end goals: execute any query as fast as possible and process as many queries simultaneously as possible without affecting individual performance. Achieving both goals is possible only if the database engine uses efficiently all CPU cores of the server and avoids bottlenecks, which cause the so-called contention during simultaneous read and write operations.

The audited SNB results demonstrate that GraphDB scales the number of read and write operations linearly from just 1 agent (3 ops/sec) all the way to 4 agents (12 ops/sec). The engine can handle in parallel multiple streams of complex graph analytical queries, while simultaneously updating the analyzed graphs in transactionally safe and consistent manner. This way the benchmark results prove that the throughput that GrahDB can handle on a single server scales up with the number of the licensed CPU cores.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com



Getting better all the time

LDBC's Semantic Publishing Benchmark was created to replicate the workload of a popular mass media, which uses a graph database to update a large number of topical web pages during a big sports event. It is based on the real case of the BBC, which successfully operated such a website with 800 pages about the FIFA World Cup in 2010 – one for each team, player, group, etc.

The BBC used GraphDB (under its former name OWLIM) to serve this website in what appeared to be the first usage of graph databases for such a high-profile critical system. The workload of such systems is about serving hundreds of queries per second, which aggregate the most relevant content for a specific topic, while at the same time handling a continuous flow of editorial updates, which should be processed instantaneously. The benchmark involves reasoning, geo-spatial constraints and full-text search, stretching the engine by making query optimization and execution really challenging.

The latest audited SPB results of GraphDB demonstrate a noticeable improvement over the previous audited results from 2015: a 6-fold improvement of the read throughput (335

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

aggregation queries now, version 55 before at scale factor 3), while handling almost 3 times more updates (26 transactions/second now, versus 10 before). Despite the new hardware, most of this improvement resulted from Ontotext's continuous efforts to optimize all aspects of the engine, so that it can handle very complex queries without requiring all data to be loaded in memory – proven with the results at scale factor 5 (SF5, 1 billion edges).

At the same time, GraphDB improved its efficiency in scaling the throughput via parallelized handling of concurrent queries given servers with CPUs having multiple cores. This is illustrated by the single-server results for 24 read agents, delivering amazing throughput of 413 queries/second at SF3 and 158 queries/second at SF5. Finally, the SPB results also demonstrate the efficiency of the new cluster architecture introduced with GraphDB 10 – a configuration of three servers gets close to tripling the read throughput at both scale factors.

Benchmarks drive the progress

The success of the previous generation relational database management systems is attributed to the heavy competition between them, which was enabled by two factors: standard query language (SQL) and proper benchmarks, developed and audited by the Transaction Processing Council (TPC).

Ontotext co-founded LDBC 10 years ago to facilitate similar progress in the field of graph database engines. Through the years it developed benchmarks, together with other partners, and participated in various activities of LDBC. The most important contribution that vendors can make is to audit benchmark results for their own engines. Ontotext does this, unlike other vendors, which make bold claims about the performance of their engines, but never publish audited results. All that it takes is commitment, solid technology, consistent proven results and a bit of courage.

Need a reliable and robust RDF graph database for your use case?

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Benchmark Results Position GraphDB As the Most Versatile Graph Database Engine

Author: Atanas Kiryakov

Ontotext's GraphDB
Give it a try today!
Download Now

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com