

2025 Graph Data Council Member Survey Report

Motivation

The Graph Data Council (GDC, formerly the Linked Data Benchmark Council) Board of Directors decided earlier this year to “take the pulse” of the organization to help align the strategic direction with member expectations. In Q1 2025, GDC members were invited to participate in confidential interviews to discuss their reasons for joining the GDC and their priorities for the coming year. Thirteen members agreed to participate. The data collection process is described in the [Data Collection](#) section. The key findings that emerged from these discussions are summarized in the [Qualitative Analysis](#) section. The anonymized notes from the individual interviews are included in the [Appendix](#).

Data Collection

Freeform interviews were used to gather information, but the following questions were used to frame the discussion around GDC mission, value proposition, priorities, and benchmarks:

- In your view, what is the GDC mission?
- How does your organization use GDC products (e.g., benchmark or graph query language specifications, benchmark results)?
- What is the GDC value proposition for your organization?
- What are your organization’s GDC priorities for 2025?
- If your organization is running LDBC¹ benchmarks, what keeps you from commissioning an audit: price, complexity, etc.?
- What changes would you like to see in existing LDBC benchmarks? What new benchmarks does the graph community need?
- Do you or your customers make purchasing decisions based on LDBC benchmark results?

Not every question was relevant or important to each participant, so it is important to note that participants were allowed to guide the discussion. Interviews ranged from 30-60 minutes.

¹ Note that LDBC is still used to denote LDBC trademarked benchmarks and when discussing the benchmarks, their results, and their data sets.

Qualitative Analysis

When the interviews were finished, qualitative coding was used to extract common themes and member objectives from the interview notes. The sample size is too small for quantitative analysis.

Value Propositions

The interviews revealed the common tangible and intangible reasons for joining the GDC. There is a desire for vendor-neutral graph query languages and benchmarks to use for product testing, but there is some disagreement among members on the relative importance of performance vs. conformance testing. When performance matters to a member (or their customers), audited LDBC benchmark results are used for external promotion to gain a competitive advantage. However, membership is just as often driven by defensive considerations. In other words, members are sensitive to LDBC benchmark results being used against them. Similarly, members are also sensitive to modification to the benchmark specifications that could put them at a competitive disadvantage.

Being part of the graph processing community is a key intangible reason for joining the GDC. Members want “a seat at the table” when graph benchmarks and query languages are discussed. The GDC gives them an opportunity to influence the discussion. Networking with other members helps them monitor the industry direction and learn the practical use-cases that drive adoption of graph technologies.

Member Priorities

The member priorities identified during these interviews will help guide the strategic direction of the GDC. Priorities tended to fall into three, sometimes overlapping, categories: new development initiatives, benchmark improvements, and graph community efforts. It is important to note that priorities vary among members and conflicts are possible.

Leveraging AI was a common theme. There was a lot of enthusiasm around developing an AI tool to convert natural language to graph queries (similar to the [Spider 1.0: Yale Semantic Parsing and Text-to-SQL Challenge](#)), though there was disagreement on the target query language. Along the same lines, developing an AI tool to optimize graph queries was also suggested. A TEXT2GRAPH task force ([work charter](#)) has already been proposed.

GQL is strategically important to many GDC members, so it is not surprising that a GQL testing and compatibility toolkit was suggested as a new development initiative. Creating a working group to get language clarifications and to make suggestions for GQL was also suggested. It could also be merged with the [GDC Extended GQL Schema \(LEX\)](#) working group.

Members would like to see the LDBC work toward unification of LPG and RDF and become more of a liaison between industry and academia.

The LDBC benchmarks are important to many members, so several benchmark-related priorities emerged from these discussions. Members would like to see the GDC gather real-world graph database use-cases, particularly those combining graphs and LLMs. Some members are interested in streaming graphs, so they would like the GDC to develop a new benchmark specification for streaming graph computation.

A few specific benchmark updates were also suggested. Members want subgraph mining and community detection algorithms added to Graphalytics. For SNB, members want a clear versioning scheme, deletion updates, and ease of use (i.e., make this benchmark easier to run). FinBench users want “future queries” (i.e., the timer starts when the data streams in, and stops when a signal is detected) on streaming data.

Though many members (or their customers) use the benchmarks, some feel that the benchmarks are too specific to particular use-cases, updated too infrequently to prevent overfitting and abuse, too difficult to run, and too expensive to audit. They would prefer that the GDC deemphasize the benchmarks and focus on data sets and synthetic data generators instead.

Appendix - Anonymized Interview Notes

The anonymized notes from the member interviews are included here. Each participant verified that no confidential information was discussed during the interviews. However, any comments that could identify a particular member have been scrubbed, even when it meant deleting useful information. The Board of Directors erred on the side of caution to preserve anonymity.

Member A

The questions listed above were used to start the conversation. The member gave two reasons for joining the GDC. First, a trusted third party recommended that they join to keep up with graph query language developments. Their main interest is the future direction of GQL. They currently support OpenCypher, but they’re moving to GQL. Second, their membership is defensive. In other words, they monitor industry graph benchmarks that could be used against them or that they could use for competitive advantage.

Their main priority for 2025 is streaming graph benchmarks. They need a benchmark in which the data are always in motion. [SAGA-Bench](#) is not suitable because it was created with a static data mindset. LDBC FinBench measures the wrong thing. FinBench times how long it takes to get an answer after a question is asked (i.e., query the present and let me know if an event occurred). In a real-world streaming application, the timer starts when the data streams in, and stops when a signal is detected (i.e., query the future and let me know when it happens). Two academic research groups are working on this problem: Riccardo Tommasini (CNRS LIRIS, France) and David Bader (NJIT, USA).

The member is not involved with the GQL language committee, but they believe a well-designed streaming graph benchmark could influence the direction of GQL. This would be an important reason to maintain their GDC membership.

Member B

The member derives several benefits from its GDC membership:

- It gives them a feel for how other companies use graph databases.
- Membership provides networking opportunities.
- The GDC helps to build the graph processing community.
- It helps them monitor and potentially influence the direction of GQL.

They are interested in graph database usage in financial scenarios. They run LDBC FinBench internally. Their customers occasionally run LDBC benchmarks.

The member specified the following priorities for GDC in 2025 (in priority order):

1. Developing a test and compatibility toolkit for GQL
2. Gathering real-world graph database use-cases (they are particularly interested in combining graphs with LLMs)
3. Add new graph analytics algorithms (e.g., subgraph mining, community detection) to LDBC Graphalytics

In addition to these priorities, this member would also like to see the GDC become a liaison between industry and academia. They want a way for members to contribute projects that benefit the larger graph community to the GDC for ongoing maintenance and governance. Finally, they would like to see the GDC contribute to something like [Spider](#) (i.e., natural language text to SQL), but for GQL.

Member C

The LDBC benchmarks, particularly SNB, are the main reason for their membership. They use SNB to test their products. They mentioned that the development of the SNB workload did not follow a clear versioning scheme, which occasionally made it difficult for them to keep track of the changes that occurred with the data. This implies that task forces need to be diligent on this by applying semantic versioning to the data sets. They are also interested in FinBench but haven't started using it yet.

The second reason for membership is the opportunity to interact with other GDC members and the graph community.

They expressed some interest in adding new algorithms (particularly community detection) to LDBC Graphalytics. Graphalytics doesn't exactly cover the part of the graph landscape that they need. They are not interested in streaming graph benchmarks at this time.

Going forward, they would like to see the GDC contribute to something like [Spider](#) (i.e., natural language text to SQL), but for graph queries. They would prefer if GDC had a balance in promoting query languages.

Member D

This member's primary motivation for joining the GDC is to be part of the graph community discussion, for the exchange of ideas, to learn about graph use-cases (particularly in generative AI), and to be part of the GQL discussion. The GDC gives them "a seat at the table." It also serves as a bridge between industry and the forward-looking ideas of academia. However, the member went on to say later that the LEX WG is the reason they joined the GDC. This working group is very important to them.

They are not using LDBC benchmarks because they are too domain-specific: "Like most application benchmarks, they're not sufficiently general." They would prefer a meta benchmark or CI/CD framework/library that takes a graph schema and query descriptions as input and generates a benchmark that reports such metrics as query runtime and memory footprint.

This year, the GDC should work to raise awareness of graph databases and consider new benchmarks for LLMs and GraphRAG.

Member E

This member's main motivation was initially graph benchmarks. Today, being part of the GDC community is their main reason for membership. They still use LDBC benchmarks for internal testing, but they commented that the benchmarks need "easier entry points" (e.g., a version of SPB that doesn't require inference or reasoning). They don't believe benchmark performance is as important to the market. The market is more interested in generative AI. They would like the GDC to extend the existing benchmarks toward AI use-cases and AI for query optimization.

Member F

In their view, the GDC mission is to produce vendor-neutral graph benchmarks and graph query languages. Influencing and monitoring GQL direction was their initial reason for joining the GDC. It is still their main priority, though they are also interested in SQL/PGQ. Most of their customers are interested in Cypher, but more of them are asking about GQL because Cypher doesn't have schema definitions.

They contend that users approach graph databases as follows. First, they check the database engine's [ranking](#), then they check publicly available benchmark results, and finally they run their own benchmarks.

They have no plans to publish audited LDBC benchmark results. They have their own internal benchmark, but they use LDBC interactive workloads and queries.

Member G

They have integrated GQL into their database product, and plan to propose new requirements for GQL this year. They believe an GDC task force or working group is needed to get clarifications and make suggestions about GQL. They suggest merging LEX with this working group.

They have not used GQL TCK for testing, but acknowledged that it will be useful in product testing.

Their customers fall into two categories: those who value LDBC audit results and those who use the LDBC datasets but design their own tests. They say there are gaps (e.g., LLM, GraphRAG, vector index) between what the LDBC benchmarks measure and what they require, so they plan to suggest changes to the LDBC benchmarks.

Member H

According to this member, the GDC mission is to facilitate the development of graph processing technologies across industries. Providing vendor-neutral benchmarks is a key value proposition. They use LDBC SNB and FinBench for internal testing. They share benchmark results with their customers. Their customers care about audited LDBC results, but the cost of audits is a barrier. They haven't decided on whether to request audited results in 2025.

Some of their customers are skeptical about the applicability to their specific query requirements. The benchmarks are sometimes too specific to a particular use-case. For example, SNB results do not necessarily translate to financial services requirements. The benchmarks and queries need to be updated to reflect new use-cases. In 2025, GDC should emphasize the benchmark data sets and data set generation. A possible solution is to design a synthetic data generator with multiple parameter configurations like the degree distribution.

GQL is also important to this member. An engine to convert text to GQL or Cypher would be useful.

Member I

All of the notes from this discussion contain potentially identifiable information. Scrubbing them removes too much context. Therefore, the notes from this discussion were redacted to preserve

member confidentiality. However, the broad points were incorporated into the qualitative analysis.

Member J

GDC's mission is to establish industry standards. Focus on developing more user scenarios, educating graph database users, and driving the growth of the graph market. The graph computing domain lacks a "killer application."

LDBC benchmarks are quite important. This member uses LDBC benchmarks to uncover system performance bottlenecks and verify ACID properties. They use the full suite of benchmarks in evaluation processes and recognize the validity of benchmark results.

SNB Interactive is well-designed, but delete updates are still needed. Running and implementing the benchmarks must also be simplified. This would increase adoption.

Users usually do not know how to use the graph technology. It would be helpful if GDC could publish more industry-specific use-cases to assist users in better adopting and using graph technology.

The GDC should work to unify LPG and RDF and to produce more user-focused cases.

Member K

Their main interest in the GDC is access to the GDC member community and SQL/PGQ discussion. A trusted third party recommended that they join the GDC to keep up with graph query language developments.

This member is interested in graph databases and implementing GQL. The LDBC benchmarks are not important to them right now, but could become important once GQL is implemented.

They believe that strong schemas are important, but a database has to work with and without them.

GQL TCK is important, but more queries need to be added. A GQL conformance suite is also needed, even though they believe that Cypher is the dominant graph query language.

Member L

The GDC value proposition for this member is as follows:

- GDC initiatives help them align with industry trends "to prevent internal overfitting."
- They use LDBC benchmarks (mainly SNB and Graphalytics) to validate system performance and competitiveness.

- They view the GDC as a forum to spread ideas.

Their current priorities are as follows:

- The GDC should expand channels for members to share findings (e.g., whitepapers, workshops, collaborative research).
- The SNB queries are useful for query optimization research, but they need more complexity and variety to allow more space for optimization. They suggest annual updates to benchmark queries to prevent vendor overfitting and to ensure system adaptability.
- They would like the cost of audits to be standardized.
- They would like to see the GDC fund open source graph projects to strengthen the GDC tooling ecosystem and community adoption.
- Publish technical blogs to demystify LDBC benchmarks and use-cases.
- Engage developer communities and publish an annual "State of Graph Systems" report to solidify GDC thought leadership.

Member M

The GDC's role as a standards organization is this member's main reason for joining. However, they are worried that the LDBC benchmarks are being abused in spite of the "fair use" provisions. They are worried about benchmark integrity, so their membership is also partly defensive.

They are interested in FinBench. They are also interested in benchmarks with real-time latency thresholds and online data ingestion and analytics. They use the LDBC datasets.

Their priorities for 2025 are as follows:

- Promote the GQL standard
- New benchmark features:
 - Add queries with real-time latency thresholds (e.g., 30-200 milliseconds)
 - Add safeguards for benchmark integrity
 Add the highly-skewed Twitter graph to SNB or some new benchmark
- There is some interest in tools to convert natural language to GQL queries