The Linked Data Benchmark Council (LDBC)

Ioan Toma[1], Irini Fundulaki[2], and Vladimir Alexiev[3]

[1] Universitaet Innsbruck
[2] Foundation for Research and Technology - Hellas
[3] Ontotext AD

1 Extended Abstract

Non-relational data management is emerging as a critical need for the new data economy based on large, distributed, heterogeneous, and complexly structured data sets. This new data management paradigm also provides an opportunity for research results to impact young innovative companies working on new RDF and graph data management technologies to start playing a significant role in this new data economy. Standards and benchmarking are two of the most important factors for the development of new information technology, yet there is still no comprehensive suite of benchmarks and benchmarking practices for RDF and graph databases, nor is there an authority for setting benchmark definitions and auditing official results. Without them, the future development and uptake of these technologies is at risk by not providing industry with clear, user-driven targets for performance and functionality.

The goal of the **Linked Data Benchmark Council (LDBC)** project is to create the first comprehensive suite of open, fair and vendor-neutral benchmarks for RDF/graph databases together with the LDBC foundation which will define processes for obtaining, auditing and publishing results. LDBC is developing benchmarks for different technology areas including *core data management* (query processing, query optimisation, transactions), *graph analysis*, *data integration* and *reasoning*. More specifically, LDBC aims at the:

- *development of new benchmarks* that will spur research and industry progress in large-scale graph and RDF data management. This includes setting challenges that will lead to significant progress in:
  - scalability, storage, indexing and query optimization techniques for RDF and graph database solutions beyond Terabyte scales.
  - quantitatively and qualitatively assess different solutions for data integration, and
  - computationally cheaper reasoning in RDF engines.
- establishment of an *industry-neutral entity*, the *LDBC foundation* for developing graph and RDF benchmarks, auditing benchmark results, and publishing audited results. The LDBC Foundation will work in the same spirit as the Transaction Processing Council (TPC) that has established a widely accepted by the industry, set of benchmarks for relational database management systems. It will be responsible for:

- specifying benchmarks, benchmarking procedures and verifying/publishing results.
- providing a TPC-style auditing service for certifying results published by vendors for benchmarks endorsed by LDBC.
- training auditors for its benchmarking, creating a long-lasting business model for auditing benchmark results.

During the EU networking session, LDBC will present the status of the two LDBC benchmarks that are currently under development:

– Semantic Publishing Benchmark (SPB)[1] which simulates the management and consumption of RDF metadata that describes media assets, or creative works. The scenario is a media organization that maintains RDF descriptions of its catalogue of creative works (e.g. from the BBC). The benchmark is designed to reflect a scenario where a large number of aggregation agents provide the heavy query workload, while at the same time a steady stream of creative work description management operations are in progress. This benchmark plainly targets RDF database systems, which support at least basic forms of semantic inference.
– Social Network Benchmark (SNB)[2] which is designed for evaluating a broad range of technologies for tackling graph data management workloads. The systems targeted are quite broad: from graph, RDF, and relational database systems to Pregel-like graph programming frameworks. The scenario of the benchmark, a social network, is chosen with the following goals in mind: it should be understandable to a large audience, and this audience should also understand the relevance of managing such data; the scenario in the benchmark should cover the complete range of interesting challenges, according to the benchmark scope; and the query challenges in it should be realistic in the sense that, though synthetic, similar data and workloads are encountered in practice.
We will also report during the EU networking session about the newly launched non-profit LDBC organization, where RDF and graph industry players and academia come together for managing the development of benchmarks as well as auditing and publishing official results.
The LDBC project is looking to network with other ongoing EU projects which are using RDF and graph databases. In particular we are looking to discuss with users of RDF technologies both researchers and industry participants about possible benchmark use cases and scenarios and to assess the quality of the current benchmark proposals and the adequacy to their needs.

---

[1] http://www.ldbc.eu:8090/display/TUC/Semantic+ Publishing+Task+Force
[2] http://www.ldbc.eu:8090/display/TUC/Social+ network+benchmark+task+force