LDBC benchmarks

Gábor Szárnyas

Graphalytics
Semantic Publishing Benchmark
Social Network Benchmark
Unofficial benchmarks
Graphalytics

**Workload:** analytical algorithms on weighted graphs

**Algorithms:** BFS, shortest paths, community detection, PageRank, etc.

**Target systems:** graph analytics frameworks

Graphalytics Competition coming up this autumn!
Semantic Publishing Benchmark

**Workload:** a semantic query workload with on a news media ontology

**Target systems:** triplestores

Complex query workload with updates (inserts/deletes)

Limited adoption so far – the SPB might be an interesting choice for **knowledge graph systems** with sophisticated inference rules
Social Network Benchmark

Workloads:

- Interactive – transactional, short-running queries, concurrent R/W
- Business Intelligence – analytical queries on daily snapshots

Target systems: DBMSs

Recent progress:

- increased adoption
- audited TuGraph, more audits planned
SNB Datagen

Key SNB component: scalable property graphs, realistic distributions, etc.

- Introduced deletions based on statistics from a defunct social network
- Migrated from Hadoop to Spark

Largest data sets generated:

- SF10,000 – 10TB CSVs, generation cost: $250
- SF30,000 – 30TB CSVs, generation cost: $800

Egress costs can get large at this scale.
We are working on hosting them in a data repository without egress costs.
Unofficial benchmarks based on the SNB

Benchmarks using SNB data sets:

- Social network analytics (SIGMOD 2014 Programming Contest)
- Labelled Subgraph Query Benchmark (GRADES-NDA 2021)

Microbenchmarks focusing on a specific computational kernel:

- Not representative of real workloads (e.g. only a few operators, no updates)
- Not suitable for system-to-system comparison
- Communicate clearly that these are not official benchmarks:
  LDBC Labelled Subgraph Query Benchmark
Summary

Existing LDBC benchmarks are used extensively in industry and academia.

The graph space has very diverse workloads – many new benchmarks are possible:

- streaming
- graph neural networks
- regular path queries

Workaround: a combination of regular benchmarks and unofficial microbenchmarks

Slides on LDBC: The Linked Data Benchmark Council