The LDBC Benchmark Suite

Gábor Szárnyas
(CWI Amsterdam, LDBC)

FOSDEM 2023 HPC room
LDBC: Linked Data Benchmark Council

- Non-profit company founded in 2012
- Designs graph benchmarks & governs their use
- Research on graph schemas and query languages
- Open-source under Apache v2

[github.com/ldbc](https://github.com/ldbc)
LDBC Social Network Benchmark

Database systems
<table>
<thead>
<tr>
<th>Data set</th>
<th>Queries</th>
<th>Updates</th>
</tr>
</thead>
</table>

Person nodes:
- Ada
- Bob
- Carl
- Dan
- Finn
- Eve
- Gia

Data set:
- Bob knows Ada

Queries:
- M1 Mon
- M2 Tue
- M3 Sun
- M4 Tue
- M5 Fri

Updates:
- author
- reply
Q9($\text{name, }$\text{day})

Data set

- Ada knows Bob
- Carlson knows Eve
- Dan knows Gia

Queries

- author
- reply

Updates

- Pa knows $*1..2$
- M
  - creation date < $\text{day}$
  - name = $\text{name}$
  - author
Q9(“Bob”, “Sat”)
Bob knows Ada
Bob knows Finn
Ada knows Finn
Bob knows Eve
Bob knows Carl
Eve knows M1
Eve knows M2
Eve knows M3
M1 Mon
M2 Tue
M3 Sun
M4 Tue
M5 Fri
M knows Pa
Pa knows Pb
Pa knows Bob
author
reply
name = “Bob”
creation date < “Sat”

Q9(“Bob”, “Sat”)
knows

+ knows("Eve", "Gia")

author

reply
Data set

Queries

Updates

+ knows(“Eve”, “Gia”)

+ Comment(“Gia”, “M3”)
+ knows("Eve", "Gia")
+ Comment("Gia", "M3")
Knows: Ada knows Finn, Finn knows Carl, Finn knows Eve, Eve knows Bob, Eve knows Dan, Dan knows Gia, Gia knows Finn, Finn knows M1, M1 knows M3, M3 knows M4, M4 knows M5, M5 knows M6, M6 knows M2, M2 knows M1.

Author: Finn

Reply: M1 Mon reply M2 Tue, M3 Sun reply M4 Tue, M4 Tue reply M5 Fri.

Updates:
- knows("Eve", "Gia")
- Comment("Gia", "M3")
Updates

- Person("Eve")
+ knows("Eve", "Gia")
+ Comment("Gia", "M3")
Updates

+ knows("Eve", "Gia")
+ Comment("Gia", "M3")
- Person("Eve")
Data set

- Person("Eve")
+ knows("Eve", "Gia")
+ Comment("Gia", "M3")

Queries

Updates

+ knows("Eve", "Gia")
+ Comment("Gia", "M3")
- Person("Eve")
Data set

- Person("Eve")

Queries

Updates

- + knows("Eve", "Gia")
- + Comment("Gia", "M3")
- - Person("Eve")
- Person("Eve")

+ knows("Eve", "Gia")

+ Comment("Gia", "M3")
Ada

Bob

Carl

Dan

Finn

M1

M2

Mon

Tue

knows

author

Cascading deletes remove lots of entities

Data set

Queries

Updates

+ knows("Eve", "Gia")

+ Comment("Gia", "M3")

- Person("Eve")

Updates

knows

Query

add lots of entities
- Runs queries and updates concurrently
- Schedules operations to be executable
- Collects results
Data set

Queries

Benchmark driver

System under test

Updates
Candidate systems:

- Graph databases
- Triplestores
- Relational databases
SELECT DISTINCT m.id
FROM ( 
    SELECT k.p2id AS id
    FROM person Pa,
    knows k
    WHERE Pa.name = $name
    AND Pa.id = k.p1id
    UNION
    SELECT k2.p2id AS id
    FROM person Pa,
    knows k1,
    knows k2
    WHERE Pa.name = $name
    AND Pa.id = k1.p1id
    AND k1.p2id = k2.p1id
    AND k1.p1id <> k2.p2id
) Pb,
Message m
WHERE Pb.id = m.authorId
AND m.creationDate < $day
SELECT DISTINCT m.id
FROM (SELECT k.p2id AS id
 FROM person Pa,
  knows k
 WHERE Pa.name = $name
  AND Pa.id = k.p1id
UNION
SELECT k2.p2id AS id
 FROM person Pa,
  knows k1,
  knows k2
 WHERE Pa.name = $name
  AND Pa.id = k1.p1id
  AND k1.p2id = k2.p1id
  AND k1.p1id <> k2.p2id
) Pb,
  Message m
WHERE Pb.id = m.authorId
  AND m.creationDate < $day
Q9($name, $day)

SELECT DISTINCT m.id
FROM (SELECT k.p2id AS id
     FROM person Pa,
     knows k
     WHERE Pa.name = $name
     AND Pa.id = k.p1id
     UNION
     SELECT k2.p2id AS id
     FROM person Pa,
     knows k1,
     knows k2
     WHERE Pa.name = $name
     AND Pa.id = k1.p1id
     AND k1.p2id = k2.p1id
     AND k1.p1id <> k2.p2id
     ) Pb,
     Message m
WHERE Pb.id = m.authorId
AND m.creationDate < $day

Graph pattern matching language with visual graph syntax inspired by Cypher

SQL/PGQ, June 2023

SELECT id FROM GRAPH_TABLE (sN
MATCH ANY ACYCLIC
(Pa:Person WHERE Pa.name = $name)
-[:knows]-{1,2} (Pb:Person)
-[:author]-> (m:Message)
WHERE m.creationDate < $day
COLUMNS (m.id))
```sql
SELECT DISTINCT m.id
FROM (
    SELECT k.p2id AS id
    FROM person Pa,
    knows k
    WHERE Pa.name = $name
    AND Pa.id = k.p1id
    UNION
    SELECT k2.p2id AS id
    FROM person Pa,
    knows k1,
    knows k2
    WHERE Pa.name = $name
    AND Pa.id = k1.p1id
    AND k1.p2id = k2.p1id
    AND k1.p1id <> k2.p2id
) Pb,
Message m
WHERE Pb.id = m.authorId
AND m.creationDate < $day
```
SNB Workloads
Queries start in 1–2 person nodes

Queries and updates run concurrently

Goal: high throughput (ops/s)

SNB Interactive

Results on the 100GB data set

Throughput (ops/s) on log scale:
- 16k
- 8k
- 4k

Year:
- 2020
- 2021
- 2022
- 2023
- 2024

Query: $Q9(name, day)$
- name = $name$
- creation date < $day$
SNB Business Intelligence

Queries touch on large portions of the data
Both bulk and concurrent updates allowed
Goal: high throughput & low query runtimes

Results on the 1TB data set

- Power@SF: 30,990
- Throughput@SF: 12,993

Diverse range of CPUs used for SNB:
- AMD EPYC Genoa
- Intel Xeon Ice Lake
- YiTian 710 (Arm v9)

- More results expected in 2023
- Graph accelerators released soon
Benchmark process

For each workload:
● Specification
● Academic paper
● Data generator
● Pre-generated data sets
● Benchmark driver
● 2+ reference implementations

Guidelines:
● How to execute the benchmark correctly
● Validate the results
● Verify ACID-compliance
Auditing and trademark

Audited benchmark runs can be conducted by independent third-party auditors

- LDBC is **trademarked** worldwide
- Only a **result produced by a certified auditor** is an “LDBC benchmark result”
- Unofficial benchmark results can be reported with a disclaimer:
  “This is **NOT** an official LDBC benchmark result”
LDBC Graphalytics Benchmark

Graph processing frameworks
(Apache Giraph, NetworKit, GraphBLAS, etc.)
Data set

- Ada
- Bob
- Dan
- Carl
- Eve
- Finn
- Gia

Untyped, unattributed graphs

Algorithms

- LDBC SNB
- Graph500
- Twitter
- Friendster
- Patents
- wiki-Talk
Data set

Algorithms

Graphalytics algorithms
Graphalytics algorithms

Breadth-first search (source: “Bob”)
Breadth-first search (source: “Bob”)
Graphalytics algorithms

Breadth-first search(source: “Bob”)
Graphalytics algorithms

Breadth-first search (source: “Bob”)
Data set

BFS: 0

BFS: 1

BFS: 1

BFS: 2

BFS: 2

BFS: 2

BFS: 3

Algorithms

Graphalytics algorithms

Breadth-first search(source: “Bob”)

PageRank(damping factor: 0.85, iterations: 5)
Graphalytics algorithms

- **Breadth-first search** (source: “Bob”)
- **PageRank** (damping factor: 0.85, iterations: 5)
Graphalytics algorithms

- Breadth-first search (source: "Bob")
- PageRank (damping factor: 0.85, iterations: 5)
- Clustering coefficient
- Community detection
- Connected components
- Shortest paths

Data set

Graphalytics spring 2023 competition – please reach out if interested
Wrapping Up...
Joining LDBC

Members can:

- Participate in benchmark design & research
- Commission audits
- Gain early access to ISO standard drafts, SQL/PGQ and GQL

Pricing:

- Free for individuals
- 2,500 EUR/year for companies
- 10,000 EUR/year for sponsor companies

Visit our website at ldbcouncil.org and reach out at info@ldbcouncil.org
**Future benchmark ideas**

**Financial Benchmark**
(to be released in 2023)

- **Target:** Distributed systems
- **Domain:** Financial fraud detection
- **Strict latency bound (20 ms)**

**Semantic Publishing Benchmark**

- **Target:** RDF/SPARQL
- **Domain:** Media/publishing industry
- **Inferencing & continuous updates**

**Graphalytics**

- **Algorithms:** BFS, CDLP, PR, SSSP, LCC, WCC
- **Data sets:** LDBC SNB, Graph500, Twitter, Friendster, Patents, wiki-Talk

---

**SNB Interactive**

- **Query:** `Q9(name, $day)`
  - `name = $name`
  - `creation date < $day`

**SNB Business Intelligence**

- **Query:** `Q11($country)`
  - `name = $country`

**Graphalytics**

- **Queries:** BFS, PR, LCC, CDLP, SSSP, WCC