



# The LDBC Social Network Benchmark (v 0.4.0) BI Workload Preliminary Result

## 14th LDBC TUC Meeting

Mingxi Wu

---

Aug 2021



# Disclaimer

- The result presented is preliminary
- It should not be used as a baseline to be compared against
- We are still optimizing some queries



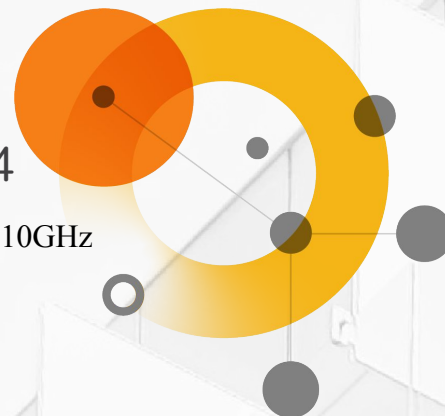
# BI Workload

- **Read Query**
  - complex read queries
  - touching a significant portion of the data.
- **Microbatches of refresh operations**
  - a set of insert and delete operations
  - batched for a given time period (e.g. a day)



# Setup - Hardware

- **Number of machines:** 24
- **Each Machine Configuration:**
  - **Operating System:** CentOS Linux 7
  - **Linux Kernel:** Linux 3.10.0-1160.el7.x86\_64
  - **CPU Model Name:** Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz
  - **Number of CPUs (threads):** 104
  - **CPU Cache:** 35.75 MB
  - **Memory:** 500G
  - **Disk:** Samsung SSD 860 EVO 2TB, Up to 500MB/s
  - **Network Bandwidth:** 9.7GB/s



# Setup - Software

- **TigerGraph RC 3.2 Enterprise Edition**



# Setup - Data

- LDBC-SNB SF-10K
- Raw Initial data: 8.9T
- Macrobatch update data:
  - Insert files: 1.3T
  - Delete files: 19G (cascade delete)

***Note: for a given batch update, say 1 day, we run insertion first, followed by deletion (since delete may include the data in the insertion).***



# Setup - Data

- Vertex delete operation triggers cascade deletion

## BI / delete / 1

|         |  |
|---------|--|
| query   | BI / delete / 1  |
| title   | Remove person and its personal forums and message (sub)threads   |
| pattern | <p>Message</p> <p>Person</p> <p>Person</p> <p>id = \$personId</p> <p>Message</p> <p>Company</p> <p>University</p> <p>Message</p> <p>likes</p> <p>knows</p> <p>workAt</p> <p>studyAt</p> <p>hasCreator</p> <p>hasInterest</p> <p>isLocatedIn</p> <p>hasMember</p> <p>hasModerator</p> <p>hasModerator</p> <p>Tag</p> <p>City</p> <p>Forum (Group/Album/Wall)</p> <p>Forum (Group)</p> <p>Forum (Album/Wall)</p> <p>⌵ invoke delete operation 6 (Posts) or operation 7 (Comments)</p> <p>⌵ invoke delete operation 4</p> |



# Setup - Data

Table. 2.2.1 Cardinality for each vertex type (total 17.69B vertices)

| Dynamic                               |                                | Static           |                                |
|---------------------------------------|--------------------------------|------------------|--------------------------------|
| Vertex Type Name <input type="text"/> | Cardinality<br>(# of vertices) | Vertex Type Name | Cardinality<br>(# of vertices) |
| comment                               | 14,092,409,518                 | company          | 1,575                          |
| post                                  | 3,378,537,355                  | university       | 6,380                          |
| forum                                 | 196,796,375                    | city             | 1,343                          |
| person                                | 20,176,525                     | country          | 111                            |
|                                       |                                | continent        | 6                              |
|                                       |                                | tag              | 16,080                         |
|                                       |                                | tagClass         | 71                             |





# Setup - Data

Table. 2.2.2 Cardinality for each edge type (total 130.82B edges)

| Edge Type Name                     | Cardinality<br>(# of edges) |
|------------------------------------|-----------------------------|
| CONTAINER_OF                       | 3,378,537,355               |
| HAS_CREATOR                        | 17,470,946,873              |
| HAS_INTEREST                       | 472,132,370                 |
| HAS_MEMBER                         | 22,003,555,186              |
| HAS_MODERATOR                      | 196,796,375                 |
| HAS_TAG                            | 24,517,772,743              |
| IS_LOCATED_IN                      | 17,491,131,353              |
| KNOWS (each edge is counted twice) | 2,772,190,554               |
| LIKES                              | 28,360,649,695              |
| REPLY_OF                           | 14,092,409,514              |
| STUDY_AT                           | 16,142,535                  |
| WORK_AT                            | 43,910,371                  |
| HAS_TYPE                           | 16,080                      |
| IS_PART_OF                         | 1,454                       |
| IS_SUBCLASS_OF                     | 70                          |



# Setup - Parameter picking

- Input parameters are generated from random values in the graph
- To guarantee an output from the query, we also write queries to explore the data first



# Setup - Cross Validation Query For Scale Factor 1

- **Validate the statistics**
  - We validated the number of vertices and edges for some of the dynamics types
- **Validate the query results**
  - The results for all 20 BI queries are cross validated per 30-day sequential updates

| SF1        | TG3.2   |         |        |        |         |         |        |          |
|------------|---------|---------|--------|--------|---------|---------|--------|----------|
| date       | Comment | nPost   | Forum  | Person | HAS_TAG | LIKES   | KNOWS  | REPLY_OF |
| 2012-09-13 | 1116485 | 999664  | 90227  | 9884   | 2471902 | 1265881 | 120835 | 1116485  |
| 2012-10-13 | 1355849 | 1055204 | 95012  | 10159  | 2820506 | 1471700 | 138487 | 1354664  |
| 2012-11-12 | 1645123 | 1111077 | 99994  | 10422  | 3194742 | 1715430 | 159946 | 1642590  |
| 2012-12-12 | 2060884 | 1168357 | 104997 | 10693  | 3710874 | 2043627 | 188572 | 2056633  |

| SF1        | Neo4j   |         |        |        |         |         |        |          |
|------------|---------|---------|--------|--------|---------|---------|--------|----------|
| date       | Comment | Post    | Forum  | Person | HAS_TAG | LIKES   | KNOWS  | REPLY_OF |
| 2012-09-13 | 1116485 | 999664  | 90227  | 9884   | 2471902 | 1265881 | 120835 | 1116485  |
| 2012-10-13 | 1355849 | 1055204 | 95012  | 10159  | 2820506 | 1471700 | 138487 | 1354664  |
| 2012-11-12 | 1645123 | 1111077 | 99994  | 10422  | 3194742 | 1715430 | 159946 | 1642590  |
| 2012-12-12 | 2060884 | 1168357 | 104997 | 10693  | 3710874 | 2043627 | 188572 | 2056633  |

# Results -Loading

- **Loading time:** 11hr43min over 24 machines
- **Loaded data size on each machine:**
  - ~190 GB (about 4.58TB in total)
- **Compress ratio:**  $8.9/4.58=0.51$
- **Loading Speed:** 32.0G/Hour/Machine



# Results - Queries with 7-day Macrobatch Update

| date                         | Comment        | Person     | LIKES          | KNOWS         | Insert time(s) | Delete time(s) |
|------------------------------|----------------|------------|----------------|---------------|----------------|----------------|
| 2012-11-29                   | 14,092,409,518 | 20,176,525 | 28,360,649,695 | 2,772,190,554 | 0              | 0              |
| 2012-12-06                   | 14,531,539,940 | 20,300,693 | 29,014,521,954 | 2,844,102,592 | 1,183          | 554            |
| 2012-12-13                   | 15,002,924,515 | 20,427,189 | 29,506,523,456 | 2,901,085,981 | 1,058          | 562            |
| 2012-12-20                   | 15,512,763,989 | 20,552,749 | 30,184,775,016 | 2,985,321,383 | 1,791          | 545            |
| 2012-12-27                   | 16,087,464,741 | 20,677,044 | 31,194,075,832 | 3,120,793,094 | 1,973          | 602            |
| <b>Total Time (s)</b>        |                |            |                |               | 6,004          | 2,262          |
| <b>Speed (GB/hr/machine)</b> |                |            |                |               | 29.93          | 1.13           |

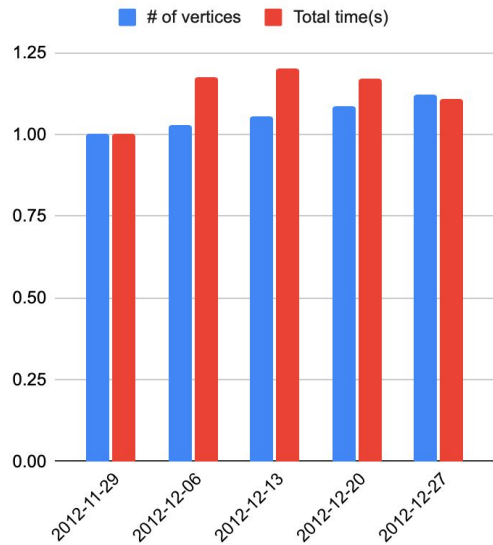
- **Insertion Speed: 29.93 G/Hour/Machine**
- **Deletion Speed: 1.13 G/Hour/Machine**
- Deletion is cascade but insertion is not. This causes deletion to be slower.

# Results - Query With Macrobatch Update

| date       | bi1    | bi2   | bi3   | bi4    | bi5   | bi6    | bi7   | bi8    | bi9   | bi10  |
|------------|--------|-------|-------|--------|-------|--------|-------|--------|-------|-------|
| 2012-11-29 | 209.15 | 10.88 | 38.68 | 67.61  | 76.81 | 386.44 | 87.48 | 135.52 | 34.32 | 12.37 |
| 2012-12-06 | 82.45  | 17.89 | 40.62 | 101.62 | 43.5  | 461.5  | 95.57 | 237.41 | 34.28 | 18.22 |
| 2012-12-13 | 72.61  | 13.13 | 43.94 | 76.86  | 42.37 | 477.44 | 97.64 | 326.57 | 39.69 | 32.57 |
| 2012-12-20 | 75.27  | 20.45 | 49.98 | 119.21 | 99.45 | 498.55 | 58.74 | 191.05 | 39.37 | 23.28 |
| 2012-12-27 | 76.15  | 16.13 | 50.12 | 100.16 | 52.26 | 543.27 | 63.23 | 206.68 | 38.25 | 14.37 |

| date       | bi11  | bi12  | bi13  | bi14   | bi15   | bi16   | bi17   | bi18 | bi19   | bi20  | Total   |
|------------|-------|-------|-------|--------|--------|--------|--------|------|--------|-------|---------|
| 2012-11-29 | 38.67 | 54.61 | 16.99 | 304.05 | 16     | 101.5  | 83     | 5.06 | 130    | 26.79 | 1835.93 |
| 2012-12-06 | 43.28 | 56.73 | 39.1  | 302.91 | 138.26 | 123.9  | 136.75 | 4.46 | 159.44 | 17.97 | 2155.86 |
| 2012-12-13 | 4.29  | 49.34 | 15.75 | 318.83 | 129.07 | 127.97 | 135.12 | 3.86 | 174.85 | 27.07 | 2208.97 |
| 2012-12-20 | 4.35  | 57.59 | 28.41 | 281.05 | 122.75 | 125.73 | 144.88 | 4.96 | 185.1  | 22.52 | 2152.69 |
| 2012-12-27 | 51.63 | 61.8  | 26.98 | 97.1   | 132.55 | 120.9  | 160.34 | 3.7  | 195.05 | 25.86 | 2036.53 |

Number of vertices and query time normalized by the initial state



- Graph size increases over time.
- The total query time increases correspondingly.

# Advertisement



Accelerate Analytics and AI with Graph Algorithms

**Stay Tuned: Virtual and In-Person Conference**

**San Francisco – October 5, 2021 | New York – October 19, 2021**