Time for a Financial Graph Database Benchmark

Xiaowei Zhu

16 August 2021

Existing Graph Benchmarks

- LinkBench

- SNB
  - Interactive
  - Business Intelligence

- Graph500

- Graphalytics

- OGB

- SPB
- BSBM
- SP2B
- LUBM
- UOBM

Why Yet Another Graph Benchmark?

- Financial **[risk control]** is an important domain for graph systems
  - Fraud detection
  - Anti-money laundering

- There are some missing aspects in existing graph benchmarks
  - Multi-edge relationships
  - Sliding window-style data management
  - Read-write operations
  - Cycles and paths with recursive temporal constraints
  - Machine learning on dynamic graphs

Financial Graph
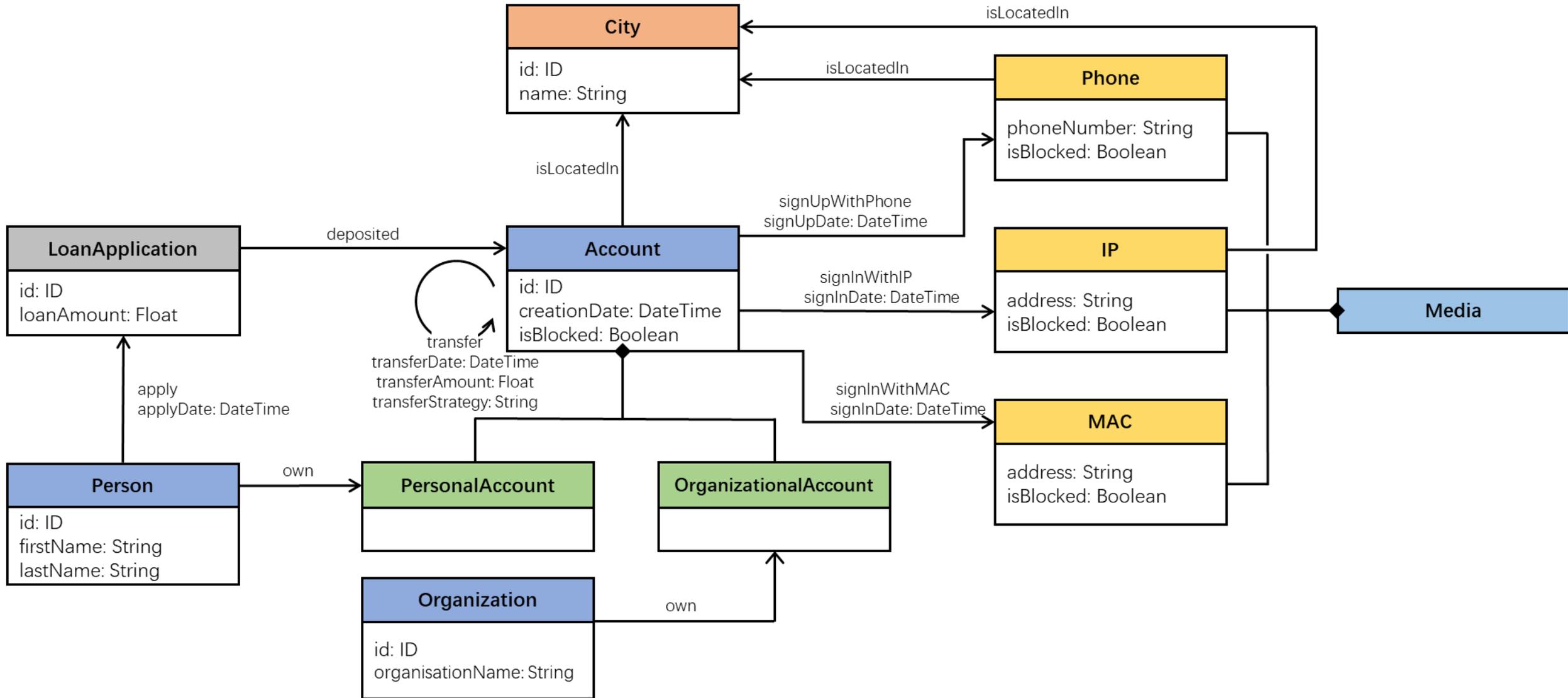
Multi-Edge Relationships

- Examples
  - Alice receives her salary once a month from her working company
  - Alice pays her rent to the landlord once a season
  - Alice bought her brunch at a cafe nearly every workday

- How to generate multi-edge synthetic property graphs?

- Why not replacing multi-edges with vertices and pairs of edges?
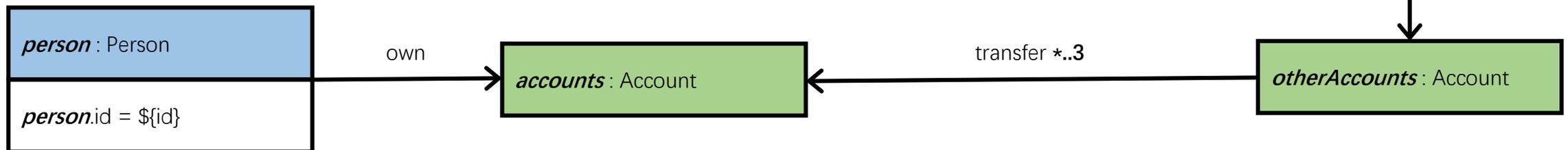
Sliding Window-Style Data Management

- Why not storing all historical data?

- Write operations
  - Upserts (most inserts) of vertices and edges with TTLs
  - Deletes of existing vertices and edges (seldom)

  - ~1:1 inserts:deletes

Read-Write Operations

- Not strictly read-write transactions
  - Read-your-writes consistency

- Examples
  - Detect whether a new edge would form new cycles
  - Count blocked accounts that are two hops away from a new vertex

Cycles and Paths with Recursive Temporal Constraints

- Given a **person**, find the **accounts** that the **person** owns.

- For each **account**, find the **otherAccounts** that the **account** receives transfers from by at most 3 steps.

- For each **otherAccount**, find the **loans** that the **otherAccount** apply for.

- Return sum of the loan amount.

- The transfers need to satisfy that the **current** edge's timestamp is within a small window of the **previous** one.

loans : LoanApplication

RESULT
SUM(**loans**.loadAmount)
COUNT(DISTINCT **loans**)

deposited

person : Person

person.id = ${id}

own

accounts : Account

transfer ***..3**

otherAccounts : Account

Machine Learning on Dynamic Graphs

- Machine learning on static graphs
  - Tensor computations
  - + Neighborhood aggregations

- Machine learning on dynamic graphs
  - + Graph updates

- How to test the performance?
  - Accuracy? Latency? Throughput?

We are looking for a full-time PI to lead the financial graph database benchmark project

Contact: robert.zxw [at] antgroup.com

Thanks!

Q & A