TOWARDS A HYBRID OLTP-OLAP GRAPH BENCHMARK

Muhammad Attahir Jibril

Technische Universität Ilmenau

Databases and Information Systems Group

The SPIRIT of science

TECHNISCHE UNIVERSITÄT ILMENAU

Storage Architecture

- **Tables:** Linked-list of chunks (multiple of 256 bytes, cache-line aligned)

- **Chunks:** fixed-sized arrays of equally-size object records, properties are outsourced

- **Record access:** 8-byte integer offsets instead of 16-byte persistent pointers

- **Lookup:** Sparse index

- **Free space management:** bitmap

- **Other data structures:** persistent dictionary for encoding property keys & values; hybrid index
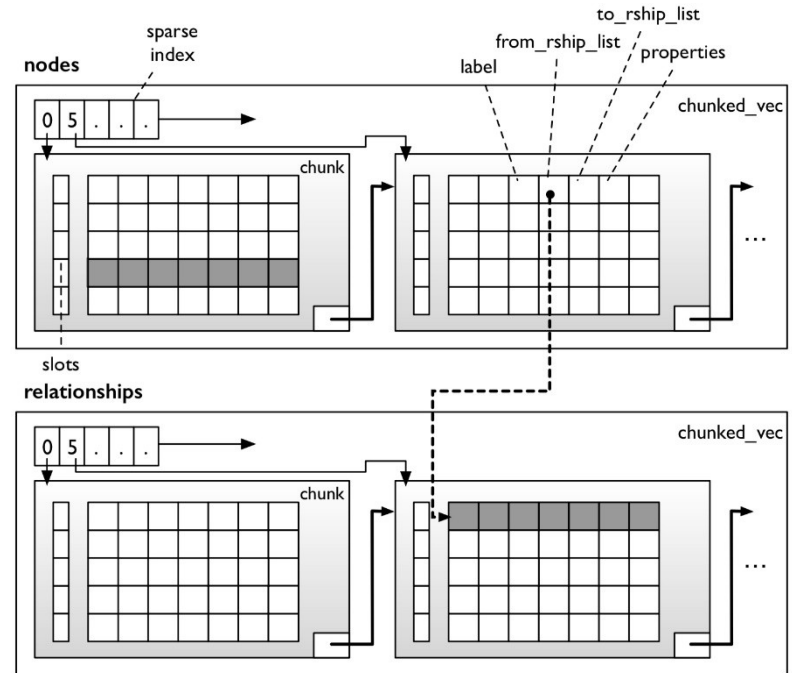
Performance Evaluation

- Evaluation of PMem, DRAM & Disk

- Scan types: sequential (s), parallel (p), indexed (i)

- LDBC-SNB Interactive Update queries (right)

- LDBC-SNB Interactive Short Read queries (bottom)
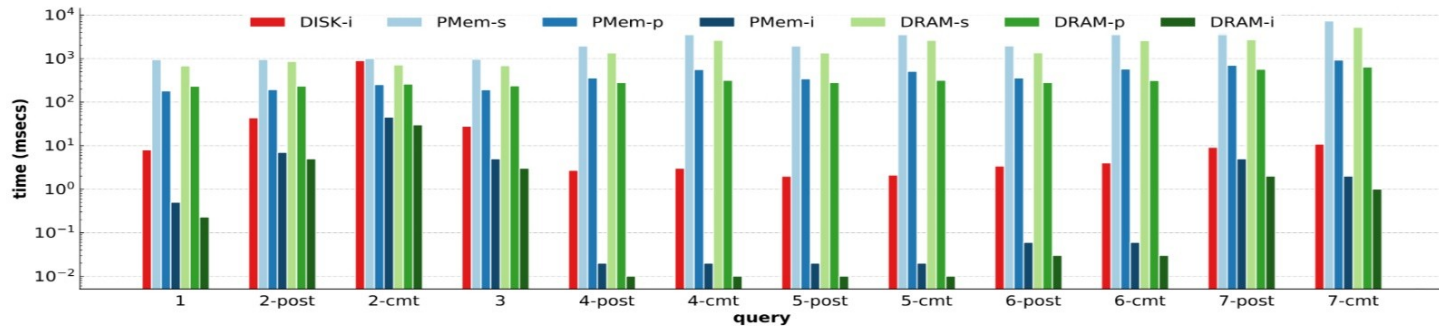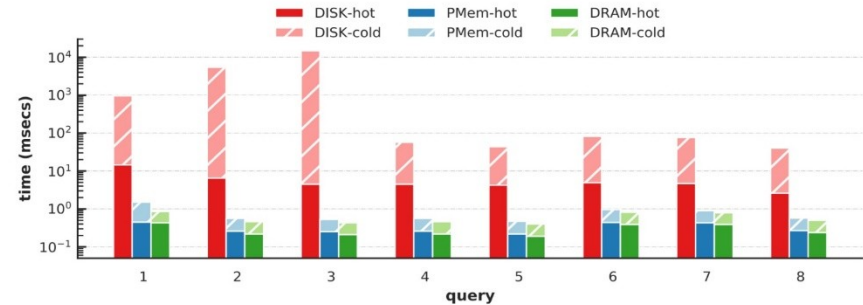
Jibril M.A. et al., JIT happens: Transactional Graph Processing in Persistent Memory meets Just-In-Time Compilation, EDBT 2021

Query Recovery

- To recover or to restart?

  - storing intermediate results, storing query states,

  - Wasted computation e.g. Aggregations & Joins

- LDBC BI queries

- Point of materialization, level of query progress

Baumstark A. et al., Instant Graph Query Recovery on Persistent Memory, DaMoN 2021 (short paper)

Hybrid OLTP-OLAP Graph Benchmark
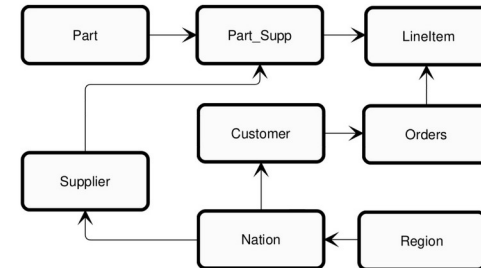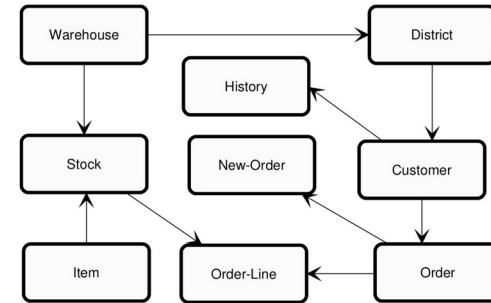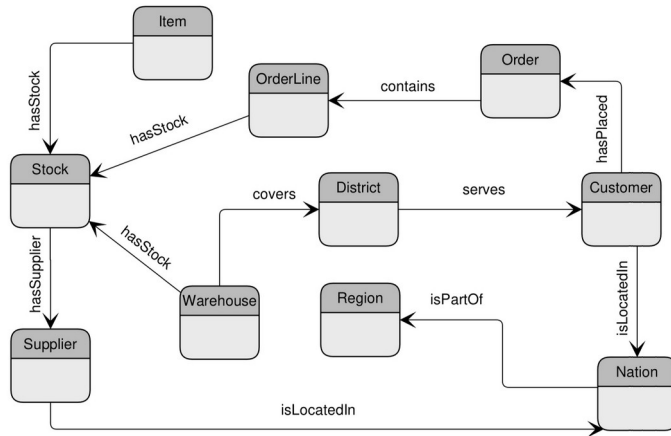
Hybrid OLTP-OLAP Graph Benchmark

- Schema Design

  - unique IDs, node vs property

  - Merge: history into customer, neworder into order

  - Common entities: customer, order

  - orderline → lineitem, item → part

 - customer-nation relationships

  - Region
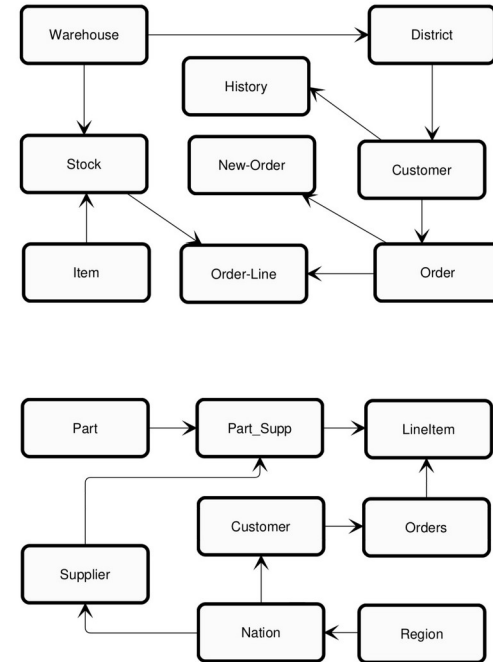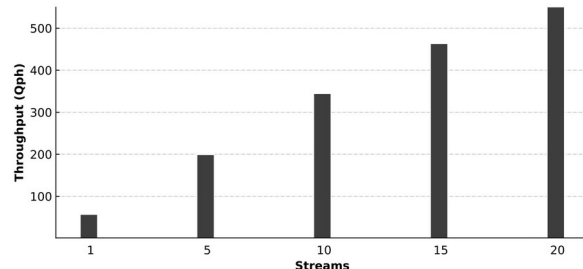
Hybrid OLTP-OLAP Graph Benchmark

Hybrid OLTP-OLAP Graph Benchmark

- Data Generation
  - relationship cardinalities
  - scaling
- Workloads
  - OLTP & OLAP queries
  - execution mode: streams
- Benchmark parameters
  - no. of warehouses & concurrent mixed streams
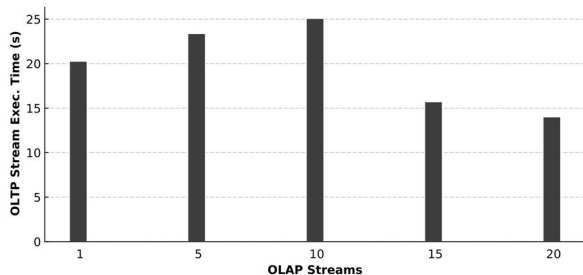- Performance metrics
  - throughput, latency

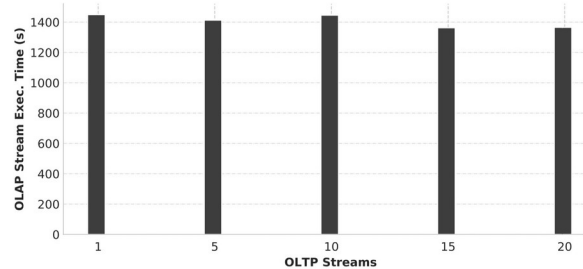# Performance Evaluation on Mixed OLTP-OLAP Graph Queries
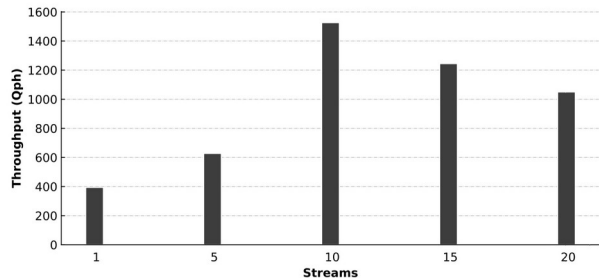
OLTP-only

OLAP-only

Varying OLAP streams

Varying OLTP streams

Outlook

- Potential schema extensions

- Data generation with more real-world characteristics

- Overall benchmark metric

Thank you