Social Network Benchmark: Virtuoso Experience

Orri Erling

Openlink Software

November 14, 2014

What is Graphiness?

- Walk of environment hits same place many times
- Paths are objects
- Cardinality estimation needs to know about transitivity
- Reuse intermediate or precompute them as hash build side

Implementation Experience

- SQL and SPARQL
- Query optimization generally easier in SQL
- Structure awareness will make SPARQL comparable to SQL
- SQL easier to analyze, more transparent

Scaling

- Larger graph have larger degree
- Lookups usually 2 steps: 10x the scale makes 2x more vertices hit
- Interactive is $log(n) \cdot avg\_degree^{1.8}$
- BI is $n \cdot log(n)$
- Graph analytics is $n \cdot log(n) \cdot degree$ onward
- 300 Gb gets ca.130 G working set (compressed column store)
- Updates are fast but need durability outside of log
- Disk write OK with 1 disk per 50 Gb, running 1.5x faster than simulation time