SNB Business Intelligence Workload
Design and Roadmap

Andrey Gubichev

TUM
with: OGL

November 14, 2014

Business Intelligence Workload for SNB

- choke points-based design
- queries touch a lot of data
- cover usual analytical challenges, plus graph-based and full-text search
- 24 query; preliminary formulation in SQL

Transitivity in BI queries

Transitivity is what makes the SNB different from relational benchmarks:

- Transitivity in dimensions (*small trees*)
  - e.g., geographical hierarchy
- Transitivity in replies to posts (*large trees*)
  - find the longest reply chain to posts in the network
- Transitivity in friendship (*graphs*):
  - find most central users in the graph

Example: BI Query 7

Most authoritative user posting on a given topic

- Query: Find 100 most authoritative users posting on a given topic.
- The authority is a number of likes received to the user's post on a given topic, such that likes from much liked posters weigh more.
- *Business Question*: expert search
- *Choke Points*: PageRank-style computation, reusing results, subqueries, index vs hash

BI Roadmap for next 6 months

- Full definitions of queries in English
- Formulations in Cypher, APIs, etc.
- Updates
- Power vs Throughput experiments
- Metrics

SQL queries:
`https://github.com/ldbc/ldbc_snb_implementations`

Graph analytics workload

- Connected components
- Single source shortest paths
- Page rank
- Community detection
- Betweenness centrality

Reference implementation in Jan 2015 (Virtuoso Open Source SQL)