

MOCHA 2017 as a Challenge for Virtuoso

Authors: Mirko Spasić, Miloš Jovanovik



MOCHA @ ESWC 2017
Portoroz, Slovenia



HOBbit
Holistic Benchmarking
of Big Linked Data



Horizon 2020, GA No 688227

May 30th, 2017

- Modern enterprise-grade solution for
 - Data access
 - Integration
 - Relational database management
 - Scalable RDF Quad Store
- SQL Tables and/or RDF based Property/Predicate Graphs
- Initially:
 - Row-wise transaction oriented RDBMS with SQL federation
 - Multi-protocol server providing ODBC/JDBC access to relational data
 - Built-in HTTP server providing a DAV repository, SOAP and WS* protocol end-points and dynamic web pages in a variety of scripting languages

- Re-targeted as an RDF graph store with built-in SPARQL and inference
- Column-wise compressed storage and vectored execution
- Virtuoso applications are in the RDF domain, with terabytes of RDF triples which do not fit into main memory
- Relational analytics
- Shared cluster capability for scale-out

- Single table of four columns - RDF QUAD
- **G** for graph, **P** for predicate, **S** for subject, **O** for object
- **P**, **G** and **S** are IRI IDs - a Virtuoso custom data type
- **O**s of string type which are longer than 12 characters are assigned a unique ID and this ID is stored as the **O** of the quad table
- Five covering indices, **PSOG**, **POSG**, **SP**, **GS**, and **OP**
- Compression implemented at two levels:
 - Within each database page, Virtuoso stores distinct values only once and eliminates common prefixes
 - Applying "gzip" to database pages
- SPARQL queries are translated into SQL at the time of parsing

- Input:
 - Duration of the benchmark: 600,000
 - Number of insert queries per stream: 100
 - Population of generated data: 10,000
 - Number of data generators - agents: 4
 - Name of mimicking algorithm: `TRANSPORT_DATA`
 - Seed for mimicking algorithm: 100
 - Number of task generators - agents: 1
- Main KPI:
 - Maximum triples per second: **1200-1300** (Virtuoso 7.2 Opensource)
 - Maximum triples per second: **5000-6000** (Virtuoso 8.0 Commercial Edition)

- Input:
 - Number of operations: 15,000
 - Scale factor: 1
- Main KPIs:
 - Loading Time: <250s (Virtuoso 7.2 Opensource)
 - Loading Time: <300s (Virtuoso 8.0 Commercial Edition)
 - Throughput: 40queries/second (Virtuoso 7.2 Opensource)
 - Throughput: 35queries/second (Virtuoso 8.0 Commercial Edition)

- Input:
 - Seed: 100
- Main KPIs:
 - Total query per second: **5queries/second** (Virtuoso 7.2 Opensource)
 - Total query per second: **5queries/second** (Virtuoso 8.0 Commercial Edition)

