

RDF Needs Better Database

- Database science is all one, laws of physics apply to all
- RDF is relational in spirit, with some extra complications
- To realize RDF's potential, it must meet the best in databasing, no separation of communities and no reinventing wheels

Virtuoso 7 Brings the Best an

Analytics DB to RDF

- Column store,, 6 to 14 bytes a quad, depending on data, 3x more compact than before
- Vectored execution Get the best of the metal, take advantage of locality 3x the thread speed over previous
- Intra-query parallelization everywhere
- Scale out with flexible deployment, more parallel



State Of Play

- Star Schema Benchmark
- Single Sever, 30G
 - Virtuoso SQL: 8.9s
 - MonetDB SQL: 18.9s
 - Virtuoso SPARQL: 45.4s
- 2 Node Cluster, 300G
 - Virtuoso SQL: 43.4s

State Of Play Cont'd ...

- Run with RDF in Virtuoso and run circles around many RDBMS.
- Use task-specific schema for performance on the level of the best in SQL analytics
- Gap between SPARQL (schema-less) and SQL (schema-first) closing, Now 5x, going to 3x, parity is reachable via self-tuning physical data layout



Big data is not only queries

- Many things, eg graph analytics cannot be expressed as queries.
- Data preparation and ETL pipeline
- Map reduce and
- Bulk Synchronous Processing (BSP) frameworks
- Virtuoso can do all that with parallel SQL procedures and Java hosting



Future convergence

- RDF is in the end about relations
- The URI, the Superkey is the ultimate silo breaker
- Get Schema-less flexibility, self describing data and schema first performance through adaptive physical layout
- Open RDF data for use by the SQL tool chain