PigSPARQL
- A SPARQL Query Processing Baseline for Big Data -

Alexander Schätzle • Martin Przyjaciel-Zablocki • Thomas Hornung • Georg Lausen

1. Motivation
- RDF datasets grow continuously in size → scalability of query processing?
- Single-place RDF stores limited in scale
- Reuse existing infrastructures and frameworks for distributed processing of Big Data
- Wide spread adoption of Hadoop MapReduce makes it an interesting candidate for distributed SPARQL processing

2. PigSPARQL
- SPARQL 1.0 engine on MapReduce for adhoc query processing of large RDF graphs
- Uses Pig (Latin), a data analysis platform on top of MapReduce, as intermediate layer between SPARQL and MapReduce
- Focus on rather costly queries involving many joins that cannot be executed in real-time at web-scale ⇒ offline processing
- Available for download *

Advantages of using Pig:
- Compatibility to future changes of Hadoop as they are covered by Pig
- Pig’s processing framework is continuously optimized and enhanced with new features

3. Architecture and Design

4. Experiments

5. Related Research & Download
- PigSPARQL: Mapping SPARQL to Pig Latin
  - SWIM 2011, in conjunction with SIGMOD 2011, Athens (Greece)
  - Cascading Map-File-Join over HBase for Scalable Join Processing
  - SSW+HPCSW 2012, in conjunction with ISWC 2012, Boston (USA)
- Map-File Join for Scalable SPARQL BGP Processing
  - IEEE CloudCom 2013, Bristol (UK)
- Large Scale RDF Processing with MapReduce
  - Book Chapter in: Data Processing Techniques in the Era of Big Data, 2014

6. Summary

Conclusion
- PigSPARQL, an implemented translation from SPARQL to Pig Latin
- Pig translates Pig Latin into MapReduce jobs and executes them in parallel on Hadoop
- It’s an easy to use and competitive baseline for the comparison of MapReduce based SPARQL processing
- With the support of SPARQL 1.0, it already exceeds functionalities of most existing research prototypes

Future Work
- Support for SPARQL 1.1 operators
- Integration of investigated join techniques

* http://dbis.informatik.uni-freiburg.de/PigSPARQL