A Linked Data reasoner in the Cloud

Jules Chevalier
LT2C, Télécom Saint-Etienne, Université Jean Monnet
jules.chevalier@univ-st-etienne.fr

Research Question

The Semantic Web helps in changing plain information into Linked Data, which contain implicit knowledge that can be leveraged using a reasoner. Reasoning is a complex process, and current solutions aim at reasoning at the scale of the Web of Data. We so need more powerful reasoners, scalable enough to make inference over very large datasets. Distributing and parallelizing this process over a cluster seems the most adapted solution. Cloud Computing appears like an interesting environment for large scale inferencing. The research question is therefore to propose a Cloud-ready Linked Data reasoner, whose architecture makes it possible to reason over a large scale corpus in a distributed way.

State of the Art

- **WebPie** [8]
  - Uses MapReduce [1] to reason over RDFS and OWL Horst [7].
  - Each rule execution is distributed.
  - WebPie suffer from issues due to MapReduce [4], despite of some upgrades [9].
- **MapResolve** [5]
  - Highlights MapReduce main problems.
  - Proposes another solution over more expressing fragments, still using MapReduce.
  - Doesn't provide significant improvements over WebPie performances.
- **Parallel inferencing for OWL knowledge bases** [6]
  - Proposes two ways for reasoning partitioning: rules splitting or data splitting.
  - The rules are splitted thanks to a rule-dependency graph partitioning.
  - Three methods for data partitioning: graph, hash and domain-specific partitioning.
  - Data are still in hermetic cores, which generates loops and duplicates.

Proposed Approach

Shared Memory

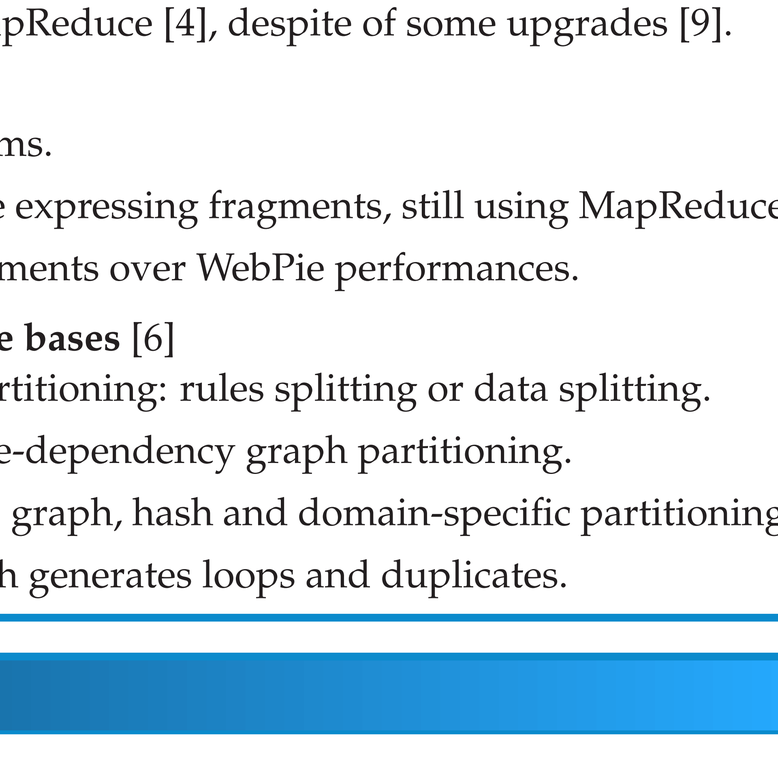
- Process memory shared by all the cores
- Each core uses the entire dataset during the process
- Cores access new triples as soon as they are inferenced
- Aims of avoiding duplicates and unnecessary loops, and speed up computation

Stream inferencing

- Use a input stream for reasoning
- Reason over new data as soon as they arrive
- Aims to avoid overhead bottle necks

Axioms sorting

- Reason over OWL-Full is computationally expensive (undecidable).
- Reasoners use less expressive fragments, but more computable.
- There are some defined fragments, like OWL Horst [7] and ρDF [3], with a good complexity/expressivity trade-off.
- We will define a parametric fragment, that can be adapted to the wanted complexity/expressivity trade-off.
- Concepts will be sorted thanks to their complexity and their rank in the figure on the right.



Rank of most used concepts usable in inference, with fragments highlighting (data from [2])

Future Work

1. Deploy WebPie and reproduce results
2. Propose and implement our stream reasoner
3. Compete against WebPie results

Schedule

- **May 2013** - State of the art internal report.
- **October 2013** - WebPie deployment and tests.
- **February 2014** - Proposal of our Cloud-hosted Linked Data reasoner.
- **May 2014** - First implementation of our reasoner.
- **November 2014** - 'Stable' version deployed on our private Cloud.
- **January 2015** - Evaluation campaign and interpretation of results.
- **April 2015** - Writing the PhD thesis.

References

[1] J. Dean and S. Ghemawat. MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.

[2] B. Glimm, A. Hogan, M. Krötzsch, and A. Polleres. OWL: Yet to arrive on the Web of Data? *LDOW*, 2012.

[3] S. Muñoz, J. Pérez, and C. Gutierrez. Minimal deductive systems for RDF. *The Semantic Web: Research and Applications*, 4519:53–67, 2007.

[4] P. Patel-Schneider. Comments on webpie. *Web Semantics: Science, Services and Agents on the World Wide Web*, 15(3), 2012.

[5] A. Schlicht and H. Stuckenschmidt. Mapresolve. *Web Reasoning and Rule Systems*, 6902:294–299, 2011.

[6] R. Soma and V. Prasanna. Parallel Inferencing for OWL Knowledge Bases. *2008 37th International Conference on Parallel Processing*, pages 75–82, 2008.

[7] H. J. ter Horst. Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *J. Web Sem.*, 3(2-3):79–115, 2005.

[8] J. Urbani. *RDFS/OWL reasoning using the MapReduce framework*. PhD thesis, Vrije Universiteit - Faculty of Sciences, 2009.

[9] J. Urbani, S. Kotoulas, J. Maassen, F. van Harmelen, and H. E. Bal. WebPIE: A Web-scale parallel inference engine using MapReduce. *J. Web Sem.*, 10:59–75, 2012.

Acknowledgement

Frédérique Laforest, Christophe Gravier and Julien Subercaze, in charge of this thesis.

OpenCloudware, funded by the French Fonds national pour la Société Numérique (FSN), and is supported by Pôles Minalogic, Systematic and SCS.