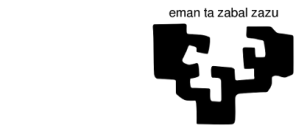
OpenLink Virtuoso as the universal database engine for the semantic web applications

Master Thesis

June 25, 2012

Łukasz Andrzej Grądzki

Supervisor:
Arantza Illarramendi, Ph.D.



Universidad del País Vasco
Euskal Herriko Unibertsitatea

Acknowledgements

This part is optional. The following are usually mentioned in the Acknowledgments:

- Supervisor and committee
- Grant support
- Helpful fellow students, lab mates etc.
- Family support

II

Preface

This thesis is submitted in complete fulfilment of the requirements for the author in the Masters's Degree in Advanced Computer Systems on The University of the Basque Country. It contains work done from May to September 2012. The supervisor of the project was Arantza Illarramendi, Ph.D. from The Facuty of Computer Science of San Sebastián. The document of this thesis has been made solely by the author who focused mainly on the analysis of the OpenLink Virtuoso database engine.

IV

Contents

List of Figures

List of Tables

Abstract

You can put an abstract of what the Thesis is about here.

Chapter 1

Introduction

1.1 Motivation

Why do we make the research? Requirements of the Semantics Web Research. Current solutions. The main research papers that are going to be used as the basis for the study. There is no definite guide for the software (feel the gap between the detailed documentation and superficial general-purpose manuals).

1.2 The OpenLink Virtuoso System

The general info about the Virtuoso System. It's properties and applications - general description.

1.3 Objectives

The main objective of the thesis[1] is to analyse and describe the properties of the Virtuoso System that are crucial for the research over the semantic webs.

Chapter 2

Problem analysis

2.1 Semantic Web

2.1.1 Definitions

Semantic Web is a project focused on definition and publication of the standards regarding content descriptions on the Internet. The main objective of these standards is to provide the content in a form convenient for the effective information processing by the software and hardware. The semantic Web standards include OWL (*Web Ontology Language*), RDF (*Resource Description Framework*, Sec. 2.2.2) and RDFS (*RDF Schema*). The meanings of the informational resources are defined using ontologies (Sec. 2.1.2) - this representation are discussed in details in the next section.

Semantic Web was an idea of Tim Berners-Lee who is the chef of W3C, the creator of the WWW standard and the first Web browser. In principle, semantic Web should be based on the existing communication protocols that set up the contemporary Internet. The main difference is that the published data must be also *comprehensible* for the machines. To achieve this objective, the resources are presented in a form that allows to identify their context and the relations between them.

2.1.2 Ontologies

In general, ontology is a formal representation of a knowledge domain, that is composed of sets of concepts and relations between them. This system creates a conceptual schema that provides a description of some domain. In addition, the conceptual schema can be used as a basis to draw conclusions about the properties of the terms described by a given ontology. According to XYZ, Ontologies are classified into lightweight ontologies and heavylight

ontologies according to it expressiveness [4]. The lightweight ontology, sometimes called terminology, is simple a taxonomic structure of concepts and some includes simple comments on relations between concepts. The heavyweight ontology extensively axiomatizes concepts and relations to represent ontological commitment explicitly and it is thus composed of concepts, relations and rules. Every heavyweight ontology can have a lightweight version. And there is not a clear borderline between light to heavy weight.

2.1.3 Field of study

2.1.4 Research trends

2.1.5 Data Storage

2.2 Data Storage Formats

2.2.1 Relational DBs

2.2.2 Resource Description Framework

2.3 Overview of existing systems

...

2.3.1 Commercial Platforms

Platform 1

...

2.3.2 OpenSource Platforms

Platform 11

...

2.4 General properties of the Virtuoso System

2.4.1 History of development

2.4.2 Properties of Virtuoso

2.4.3 Main applications

2.4.4 Aspects important for the Semantic Webs

2.5 Used Tools

2.5.1 LUBM Benchmark

The Lehigh University Benchmark is developed to facilitate the evaluation of Semantic Web repositories in a standard and systematic way. The benchmark is intended to evaluate the performance of those repositories with respect to extensional queries over a large data set that commits to a single realistic ontology. It consists of a university domain ontology, customizable and repeatable synthetic data, a set of test queries, and several performance metrics.

Chapter 3

System analysis

3.1 Internal Structure

3.1.1 SPARQL end-point

3.1.2 Advantages

3.1.3 Special queries and commands

3.2 Key Features

3.2.1 Reasoning

Forward/backward chaining.

3.2.2 Import Data Mechanisms

3.2.3 Export Data Mechanisms

3.3 Comparison with similar systems

3.3.1 Virtuoso vs Oracle NoSQL

Chapter 4

Conclusions

The section includes the following: Overall analysis and integration of the research and conclusions of the thesis in light of current research in the field Conclusions regarding goals or hypotheses of the thesis that were presented in the Introduction, and the overall significance and contribution of the thesis research Comments on strengths and limitations of the thesis research Discussion of any potential applications of the research findings An analysis of possible future research directions in the field drawing on the work of the thesis ...

Appendix I

First appendix

Contents

This is the first appendix.

I.a First section

This is the first section of the first appendix.

References

[1] S. Wind, "Open source cloud computing management platforms: Introduction, comparison, and recommendations for implementation," in *Open Systems (ICOS), 2011 IEEE Conference on*, pp. 175 –179, sept. 2011.

Publications

This is a list of publications.

- **My first article**
  M. Y. Name
  Journal **year**, *volume*, pages.