Accepted Manuscript

Ontology Based Data Access in Statoil

Evgeny Kharlamov, Dag Hovland, Martin G. Skjæveland, Dimitris Bilidas, Ernesto Jiménez-Ruiz, Guohui Xiao, Ahmet Soylu, Davide Lanti, Martin Rezk, Dmitriy Zheleznyakov, Martin Giese, Hallstein Lie, Yannis Ioannidis, Yannis Kotidis, Manolis Koubarakis, Arild Waaler

PII: S1570-8268(17)30027-6
DOI: http://dx.doi.org/10.1016/j.websem.2017.05.005
Reference: WEBSEM 437

To appear in: *Web Semantics: Science, Services and Agents on the World Wide Web*

Received date : 26 October 2016
Revised date : 14 March 2017
Accepted date : 25 May 2017

Please cite this article as: E. Kharlamov, D. Hovland, M.G. Skjæveland, D. Bilidas, E. Jiménez-Ruiz, G. Xiao, A. Soylu, D. Lanti, M. Rezk, D. Zheleznyakov, M. Giese, H. Lie, Y. Ioannidis, Y. Kotidis, M. Koubarakis, A. Waaler, Ontology Based Data Access in Statoil, *Web Semantics: Science, Services and Agents on the World Wide Web* (2017), http://dx.doi.org/10.1016/j.websem.2017.05.005

Ontology Based Data Access in Statoil

Evgeny Kharlamov[a], Dag Hovland[c], Martin G. Skjæveland[c], Dimitris Bilidas[b], Ernesto Jiménez-Ruiz[c], Guohui Xiao[d], Ahmet Soylu[f], Davide Lanti[d], Martin Rezk[d], Dmitriy Zheleznyakov[a], Martin Giese[c], Hallstein Lie[e], Yannis Ioannidis[b], Yannis Kotidis[g], Manolis Koubarakis[b], Arild Waaler[c]

[a] *University of Oxford, Department of Computer Science, Wolfson Building, Parks Road, OX1 3QD, Oxford, UK.*
[b] *National and Kapodistrian University of Athens, Panepistimiopolis, Ilissia, 15784, Athens, Greece.*
[c] *Department of Informatics, University of Oslo, Blindern, 0316, Oslo, Norway.*
[d] *KRDB Center, Free-University of Bozen-Bolzano, P.zza Domenicani 3, 39100, Bolzano, Italy.*
[e] *Statoil ASA, Stavanger, Norway.*
[f] *NTNU – Norwegian University of Science and Technology, Teknologiveien 22, 2815, Gjøvik, Norway.*
[g] *Athens University of Economics and Business, 76 Patission Street, 10434, Athens, Greece.*

Abstract

Ontology Based Data Access (OBDA) is a prominent approach to query databases which uses an ontology to expose data in a conceptually clear manner by abstracting away from the technical schema-level details of the underlying data. The ontology is 'connected' to the data via mappings that allow to automatically translate queries posed over the ontology into data-level queries that can be executed by the underlying database management system. Despite a lot of attention from the research community, there are still few instances of real world industrial use of OBDA systems. In this work we present data access challenges in the data-intensive petroleum company Statoil and our experience in addressing these challenges with OBDA technology. In particular, we have developed a deployment module to create ontologies and mappings from relational databases in a semi-automatic fashion; a query processing module to perform and optimise the process of translating ontological queries into data queries and their execution over either a single DB of federated DBs; and a query formulation module to support query construction for engineers with a limited IT background. Our modules have been integrated in one OBDA system, deployed at Statoil, integrated with Statoil's infrastructure, and evaluated with Statoil's engineers and data.

Keywords: Ontology Based Data Access, Statoil, Optique Platform, System Deployment, Evaluation, Bootstrapping, Optimisations.

1. Introduction

The competitiveness of modern enterprises heavily depends on their ability to make the *right* business decisions by relying on efficient and timely analyses of the *right* business critical data. For example, one of the factors determining the competitiveness of Statoil[1,2] is the ability of its exploration geologists to find in a timely manner new exploitable accumulations of oil or gas in given areas by analysing data about these areas. Gathering such data is not a trivial task in Statoil and in general in data intensive enterprises due to the growing size and complexity of corporate information sources. Such data sources are often scattered across heterogeneous and autonomously evolving systems or has been adapted over the years to the needs of the applications they serve. This often leads to the situation where naming conventions for schema elements, constraints, and the structure of database schemata are very complex and documentation may be limited or nonexistent making it difficult to extract data. As a result, the data gathering task is often the most time-consuming part of the decision making process.

Indeed, Statoil geologists often require data that is stored in multiple complex and large data sources that include EPDS, Recall, CoreDB, GeoChemDB, OpenWorks, Compass, and NPD FactPages (see Section 2 for details about these DBs). These DBs are mostly Statoil's corporate data stores for exploration and production data and Statoil's interpretations of this data. Some of these

Email addresses: evgeny.kharlamov@cs.ox.ac.uk (Evgeny Kharlamov), hovland@ifi.uio.no (Dag Hovland), martige@ifi.uio.no (Martin G. Skjæveland), d.bilidas@di.uoa.gr (Dimitris Bilidas), ernestoj@ifi.uio.no (Ernesto Jiménez-Ruiz), xiao@inf.unibz.it (Guohui Xiao), ahmet.soylu@ntnu.no (Ahmet Soylu), davide.lanti@unibz.it (Davide Lanti), mrezk@inf.unibz.it (Martin Rezk), dmitriy.zheleznyakov@cs.ox.ac.uk (Dmitriy Zheleznyakov), martingi@ifi.uio.no (Martin Giese), hali@statoil.com (Hallstein Lie), yannis@di.uoa.gr (Yannis Ioannidis), kotidis@aueb.gr (Yannis Kotidis), koubarak@di.uoa.gr (Manolis Koubarakis), arild@ifi.uio.no (Arild Waaler)

[1] Statoil ASA, is a Norwegian multinational oil and gas company headquartered in Stavanger, Norway. It is a fully integrated petroleum company with operations in thirty-six countries.

[2] https://www.statoil.com/