School of Systems Engineering

University of Reading

Investigating Elastic Cloud Based Reasoning for the Semantic Web

Omer Dawelbeit

Introduction

The Semantic Web [1] is an extension of the traditional Web whereby data can have a meaning. The vision of the Semantic Web is to enable intelligent agents to understand and reason over the data on the Web in order to infer new knowledge. This knowledge can help humans with many tasks such as question answering, intelligent semantic search, social discovery, etc…

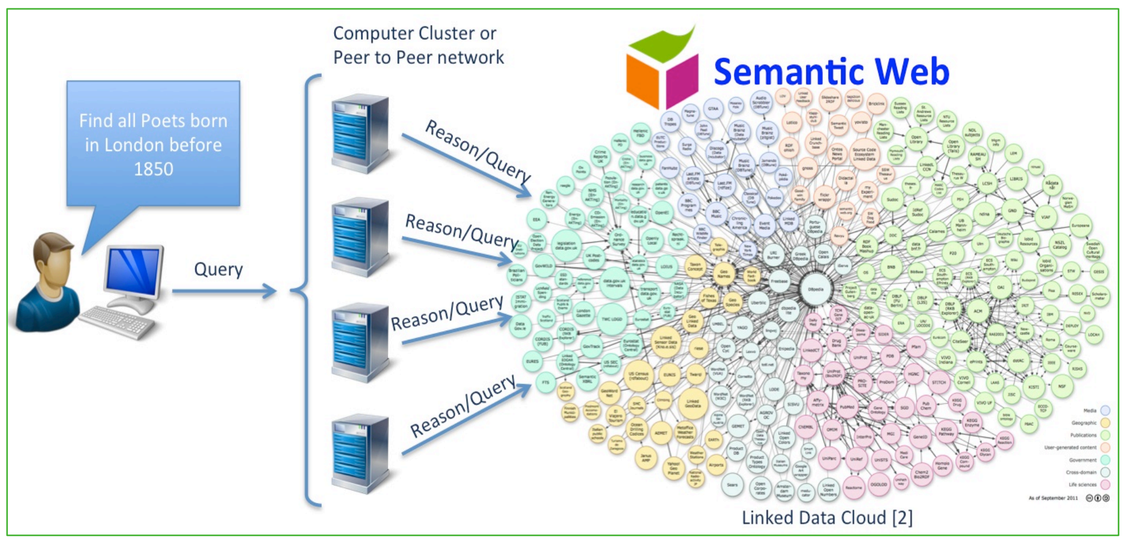

Figure 1. A user querying massive Semantic Web linked data.

Intelligent agents reasoning over large scale Semantic Web data in the region of billions of statements, need to be scalable and efficient. The computational power required far exceed the capabilities of one computer, hence the algorithms used need to distribute the processing to a large number of computers (Figure 1).

Challenges arise to due to data interdependence and distribution skew. So two questions arise:

1. **How can we divide the reasoning tasks fairly between a number of computers and merge the results?**
2. **How can we carry out this process in a cost efficient manner?.**

Research Aim & Objectives

The primary aim of the research is to develop a novel **Elastic, Cost Aware Reasoning Framework (ECARF)** for Semantic Web reasoning and evaluate it through the development of a prototype.

Objectives

The main objectives of this research are outlined as follows:

- To review the theories of distributed large scale Semantic Web reasoning algorithms.
- To analyse these algorithms and investigate how to apply them to cloud computing.
- To identify factors impacting the cost of cloud based Semantic Web reasoning.
- To develop a novel framework for elastic, cost aware cloud based reasoning for the Semantic Web (ECARF). This will be the contribution of this research.
- To evaluate the framework through a cloud based prototype.

Methodology

The research will follow a quantitative methodology and will be conducted in four stages:

1. **Literature review** – Conduct an in-depth literature review of the current state of the art.
2. **Theory development** – Develop the theoretical background for ECARF.
3. **Prototype development** – Develop a prototype to facilitate the evaluation of ECARF.
4. **Evaluation** – Empirical evaluation of ECARF using both real world and synthetic Semantic Web data and comparison against existing models.

Preliminary Findings

Mapping from fixed to Cloud computing

The following mappings from fixed to Cloud computing have been developed during the literature review and theory development stages:

- Distributed NoSQL Datastores (DNDs) can be used for shared storage.
- Cloud Virtual Machines (VMs) can be used to replace physical computing nodes. VMs are divided into long and short lived.
- Recursion Theorem in Computability Theory can be applied to cloud VMs (**Figure 2**).
- Consistent hash function can be used for data partitioning between VMs (**Figure 3**).

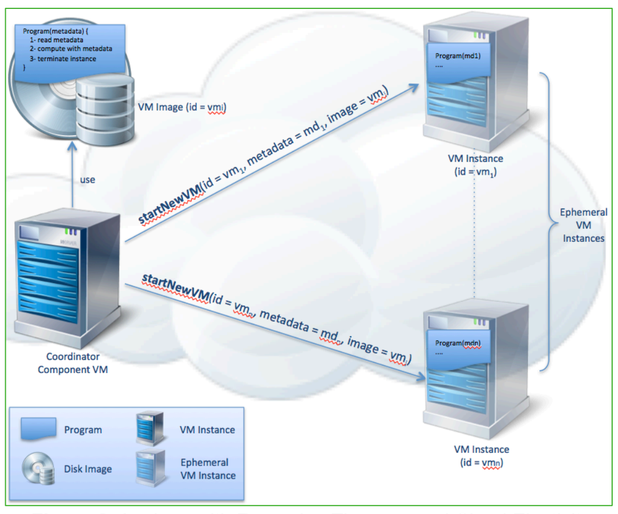

Figure 2. Analogy with Recursion Theorem and Cloud Elasticity.

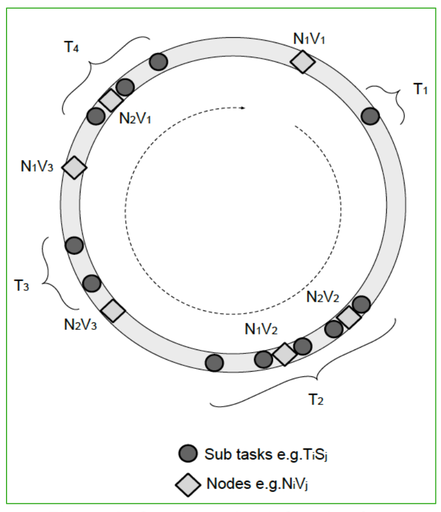

Figure 3. Consistent hash function for workload distribution.

Initial ECARF model

Definition 1. *ECARF is a 8-tuple, $(\Theta, \Sigma, T, \Gamma, \Delta, \rho, f, \sigma)$, where $\Theta, \Sigma, T, \Gamma$ are all finite sets and:*

1. Θ is the set of coordinator components VMs,
2. Σ is the set of EVMs,
3. T is the set of Tasks,
4. Γ is the set of input to be processed,
5. Δ is a DND,
6. ρ is the program that is embedded on a VM image and able to read the VM metadata,
7. $f : 2^{\Sigma} \times T \rightarrow \Sigma$, is the workload consistent hash function, and
8. σ is the analytical cost function.

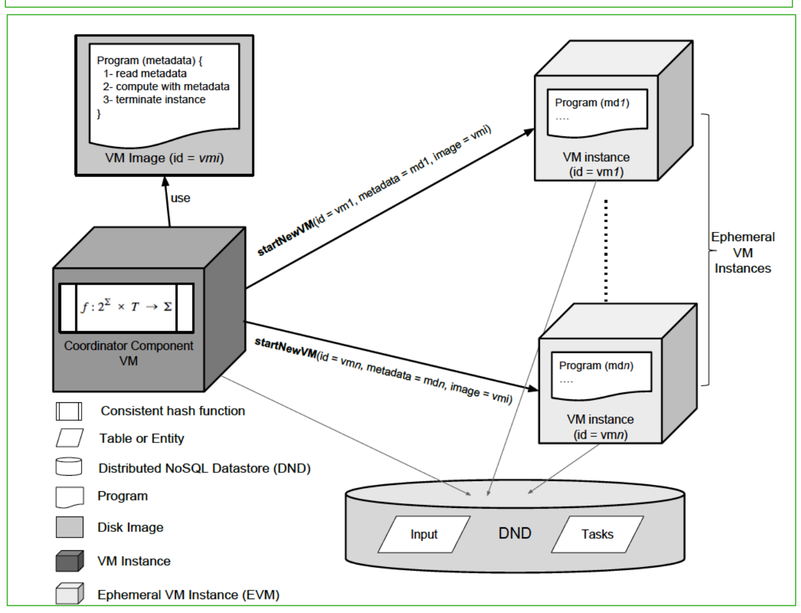

Figure 4. Initial ECARF Model.

Remaining Work

- Theoretical development of forward and backward reasoning algorithms.
- Theoretical development of Ontology management tasks (update/delete).
- Development of the prototype and deployment to Amazon Web Services.
- Evaluation of ECARF.

References

1. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific American, vol. 284, no. 5, pp. 34–43, 2001.
2. R. Cyganiak, A. Jentzsch, "Linking Open Data cloud diagram". Available online at: http://lod-cloud.net/. Last visited 12/03/2014
3. Amazon Web Services. http://aws.amazon.com/. Last visited 12/03/2014

Contact information

- School of Systems Engineering, University of Reading, Whiteknights, RG6 6AH
- Email: **o.i.o.dawelbeit@pgr.reading.ac.uk**