Thesis Advancement Report 2014-2015 (First Year)

Thesis title: Trusted-SLA Guided on Multi-cloud Environments
PhD. student: Daniel Aguiar da Silva Carvalho
Supervisor: Chirine Ghedira-Guegan
Co-supervisors: Nadia Benani and Genoveva Vargas-Solar

Context

Cloud computing is an emerging paradigm which provides on-demand computing resources in a pay-as-you-go model. In this context, we have two roles - the *cloud provider* and *cloud customer* - which have different requirements and obligations. In this model, a common way of specifying these requirements and obligations is through service level agreement (SLA) contracts agreed between the parts. Naturally, as applications and customer requirements becomes more stringent, one single cloud can be out of resources while delivering and providing all customer's desires. To avoid this, collaborations between clouds (muli-cloud) in order to share resources takes place.

The possibility of having unlimited access to resources in this multi-cloud environment opens new challenges for data processing. For instance, the data integration problem that consists in merging data from different sources and providing a unified view of the data could take advantages from this environment.

Considering the aforementioned, we are interested in proposing data integration approach in a multi-cloud hybrid environment guided by user preferences and service level agreements exported by different clouds. To the best of our knowledge, we have not identified any other proposal concerning the use of SLAs combined with a data integration approach on a multi-cloud context.

Let us show an example from the domain of energy management to illustrate our problem. So for instance, we assume we are interested in queries like: *Give a list of energy providers that can provision 1000 KW-h, in the next 10 seconds, that are close to my city, with a cost of 0,50 Euro/KW-h and that are labeled as green?* The question is how can the user efficiently obtain results for her queries such that they meet her QoS requirements, they respect her subscribed contracts with the involved cloud provider(s) and such that they do not neglect services contracts? Particularly, for queries that call several services deployed on different clouds.

Synthesis of the research activities 2014-2015

During the first year of PhD project, we have been working on the state of the art. The idea is to be aware of all types of publications close related to the thesis proposal. To reach this, we proceeded with a literature analysis using a systematic mapping methodology.

Briefly, the methodology consists in retrieving papers from scientific databases using the same search string. These papers are filtered according to an inclusion and exclusion criteria that should be defined based on the research interests. The papers will be classified in different categories (called facets) and for each facet in a specific dimension. The facets and dimension are defined based on the authors' knowledge and interests. Taking the final papers collection, the abstracts should be read in order to classify each paper into the dimensions for each facet.

The final objective of this first work is to show the research trends of data integration as a result of the emergence of the cloud and the characteristics associated to Big Data that require resources in order to be processed. In other terms, this methodology allowed us to identify trends and open issues regarding our research topic and proposing an approach that fills some gaps and proposes an original data integration solution according to current trends in the area.

This work performed, we thus have written an article that was **submitted** and **accepted** to the 26th International Conference on Database and Expert Systems applications (DEXA 2015) in the beginning of March 2015. This work has been made in collaboration with the Informatics' Lab in Grenoble (co-supervisor Genoveva Vargas-Solar) and LIRIS Lab at INSA (co-supervisor Nadia Benani). We retrieved 1832 papers. 114 papers were selected based on our inclusion and exclusion criteria. This final data collection builds the state of the art to the thesis and, in the next step, we will perform an in-depth analysis of theses papers in order to (i) analyze what have been made and (ii) propose my approach.

Perspectives for the next year 2015-2016

Based on the publications extracted from the mapping process methodology, we will proceed the analysis of the current state of the art in order to formalize our proposal. The analysis will be the basis to our model proposal. As a natural result, we will write a paper describing our approach and a survey. In parallel, we will carry on the first steps of implementation of the proposed approach. The table below describes our intended calendar. The following activities are:

1. Paper submission: systematic mapping analysis.
2. Analysis of the current state of the art.
3. Identifying the list of SLA measures associated to data used by contracts.
4. Proposal of SLA model(s) for data integration.
5. Environment configuration and deployment of services using Open Stack.
6. Description and implementation of match-retrieving algorithm.
7. Evaluation and tests of our implementation.
8. Writing a report about the experiment.
9. Approach proposal.
10. Writing the paper that describes our approach and experiment results.
11. Writing the survey.

-	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan
1	•										
2	•	•	•	•	•						
3			•								
4			•								
5			•	•							
6				•							
7				•							
8					•						
9					•	•					
10					•	•	•				
11							•	•	•		