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

The data integration is a well-known and widely discussed problem in the database area. It consists in merging data from different data sources and granting a unified view of the data[1]. The cloud computing opens new challenges to data integration. The possibility of an unlimited access to resources that arises with the cloud model changes the way to process data. In this context, some data integration approaches have been proposed. The most addressed issue in these approaches is privacy, among others such as data protection, integrity and confidentiality.

Applications and customer requirements naturally becomes more stringent. At this point, one single cloud can be out of resources while delivering and providing all requirements. In order to avoid this, cloud providers began to share their computing resources. This cloud collaboration is mentioned in the literature as multi-cloud, inter-cloud, cloud of clouds, hybrid cloud or federated cloud. The new configuration of the cloud add more challenges to data processing, considering the large amount and diversity of data, and quality and security aspects of the integration.

In the cloud model, a common way of defining requirements and obligations between the *cloud provider* and *cloud customer* is through service level agreement (SLA) contracts. SLAs have been widely adopted in the cloud context. The proposal contributions are divided in two groups: (i) approaches focus on the SLA negotiation phase; and (ii) approaches for monitoring and allocation of resources in order to detect and avoid SLA violations. Among them, we identified one approach regarding data integration in a grid environment guided by an SLA model[2].

Based on the aforementioned, we are interested in proposing a 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. This new approach brings different challenges and open issues: (i) which quality measures associated to data (that the data consumers what to query) and to the multi-cloud context should be present in SLAs? (ii) Considering the nature of the multi-cloud environment, how can quality and security aspects be guaranteed in the integration process? (iii) During the sources integration (services), it is necessary to monitor the chain of SLAs associated to different delivery models (IaaS, PaaS, etc) in order to guarantee no SLA violation in the others cloud subscriptions; and (iv) it is necessary the design of new matching-retrieving algorithms in order to select the best service composition according to the user requirements and the SLAs.

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

[1] Maurizio Lenzerini. Data integration: A theoretical perspective. In Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '02, pages 233–246, NY, USA, 2002.

[2] Tiezheng Nie, Guangqi Wang, Derong Shen,Meifang Li,and Ge Yu. SLA-based data integration on database grids. In Computer Software and Applications Conference, 2007.

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. 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. 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:

	First year								6 months 2nd year					
	1	2	3	4	5	6	7	8	9	10	11	12	1	2
1. Building a corpus using the systematic analysis method														
1.1. List of papers to read and acquire a background knowledge														
1.2 Systematic mapping														
1.3. State of the art and open issues														
2. SLA based data integration														
2.1 Analysis of existing SLA contracts and guidelines														
2.2 SLA model for data integration														
2.3 Match retrieving algorithm based on SLA														
3. Experiments and tests														
3.2 Cloud platform configuration														
3.2. Generation of synthetic services														
3.3 Implementation of the match retrieving algorithm														
3.4 Experimental tests according to a queries taxonomy														
4. PhD. Proposal														
4.1 SLA oriented rewriting algorithm for integrating data services in a multi-cloud setting														
4.2 Validation scenario