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 Bennani and Genoveva Vargas-Solar

1 Context

The data integration is a well-known and widely studied problem in the database domain. It consists in merging data from different data sources and granting a unified view [8]. The main contributions of the area are: (i) providing a global integrated representation of heterogeneous data by defining a schema (e.g., global and local as view approaches), tagging data with meta-data or by associating them to knowledge (e.g. semantic Web approaches); and (ii) architectures used for integrating data (i.e. distributed databases, multi-databases, federated databases, etc). The emergence of cloud computing and service oriented computing opens new challenges to data integration. The possibility of an unlimited access to resources changes the problems associated to data processing. Cloud-based data management using data sharing enables the collaboration of different entities to perform design tasks [6, 7]. Data processing and analytics are costly tasks that can profit from the cloud elastic resources provision, coupled with programming paradigms like Map-Reduce.[5] proposes SODIM that works on a pool of collaborative services and can process a large number of databases represented as web services.

In the cloud scenario, one cloud cannot be expected to provide the necessary resources to fulfill application requirements. Therefore, applications use different cloud providers for externalizing different data processing and management resources adding more challenges to data integration, considering the large amount and diversity of data, and quality and security aspects of the integration [4].

In cloud computing, a common way of defining requirements and obligations between the *provider* and *customer* is through service level agreement (SLA). SLAs have been adopted in the cloud, focussing on the lifecycle of a security SLA on hybrid clouds [3], on SLA models for addressing management capabilities as a service, Pcloud services, (elasticity, high availability, scalability and on demand provisioning) performed through agreed and negotiated in contracts [2]; and for addressing functional and non-functional requirements of the different cloud delivery models [1]. Summarizing, SLA contributions focus on: (i) the SLA negotiation phase; and (ii) resources monitoring and allocation to detect and avoid SLA violations. We identified one single approach regarding data integration in a grid environment guided by SLA [9].

We believe that data integration on multi-cloud environments can take advantage by integrating SLA on its solutions. To the best of our knowledge, we have not identified any other proposal adopting the use of SLAs combined with a data integration approach on a (multi)-cloud context.

2 Problem Statement

The problem addressed in our work is how can a user efficiently obtain results for her queries, meeting her QoS requirements, respecting her subscribed contracts with the involved cloud provider(s), and without neglecting services contracts? Particularly, for queries that call several services deployed on different clouds. *Hypothesis:* In order to address the problem we assume that data are provided as services that export APIs with methods to retrieve and process data. Data integration is done (i) on a (multi)-cloud service oriented environment; (ii) under new conditions with respect to the type of data sources, the environment where it is performed and the preferences of data consumers and the SLA. We assume also that (iii) SLA measures can be monitored and negotiated in all cloud providers; (iv) cloud services and data services are listed in a registry.

3 Objectives

The objective is to propose a data integration solution in a multi-cloud environment guided by user preferences and SLA exported by different clouds. This new approach brings different challenges and open issues: (1) Identify and classify quality measures associated to data and to cloud resources; (2) Propose

and implement a mechanism that ensures SLA within the data integration process performed on a multi-cloud and cope this with application requirements; and (3) Design a new matching-retrieving algorithm to perform the integration process, selecting the best service composition according to the user requirements and the SLAs.

4 Synthesis and Perspectives of the Research Activities

During the fist year we have organized our research activities in three groups (see below). These activities were organized and discussed in meetings with advisors and with individual work.

Problem statement and state of the art. The objective has been to acquire background knowledge on data integration, cloud and SLA building a corpus with publications and reading selected papers. Therefore, we applied the systematic mapping methodology that consists in retrieving papers from scientific databases, filtering them according to inclusion and exclusion criteria and research interests expressed in research questions. A classification schema must be proposed, consisting in facets and dimensions. The abstracts of the final papers collection should be read to classify each paper according to the classification scheme. We identified the trends and open in our research topic and proposed the general lines of an original data integration solution according to current trends in the area.

Results: We built a a collection of 114 and analytics results shared in (git). We proposed a data integration classification scheme that serves as initial entry for building a state of the art.

Experimentation. We are currently configuring an experimentation platform on the cloud [1].

Publications and thematic schools.

D. A. S. Carvalho, P. A. Souza Neto, G. Vargas-Solar, N. Bennani, C. Ghedira, Can Data Integration Quality be Enhanced on Multi-cloud using SLA?, Short paper, In Proceedings of the 26th International Conference on Database and Expert Systems Applications, LNCS, Valencia, Spain, 2015 (to appear)

Additionally, in April, I attended to the *1st French Brazilian School on Smart cities and Big Data* at the University of Grenoble Alpes.

The figure below presents the perspectives described as activities in the following calendar.

---

[1]You can check the detailed list of activities in https://www.dropbox.com/s/2cf6gncumzrjacd/sla-matching-experiment.docx?dl=0

	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											■	■	■	■

Milestones (diagonal labels):
Paper submission to DEXA · Accepted Paper DEXA · UML SLA model design platform · Designed and Implemented algorithm · Experimental tests and results · Paper submission to TKDLS · Proposal and scenario specification · Experiments · Scientific report or position paper
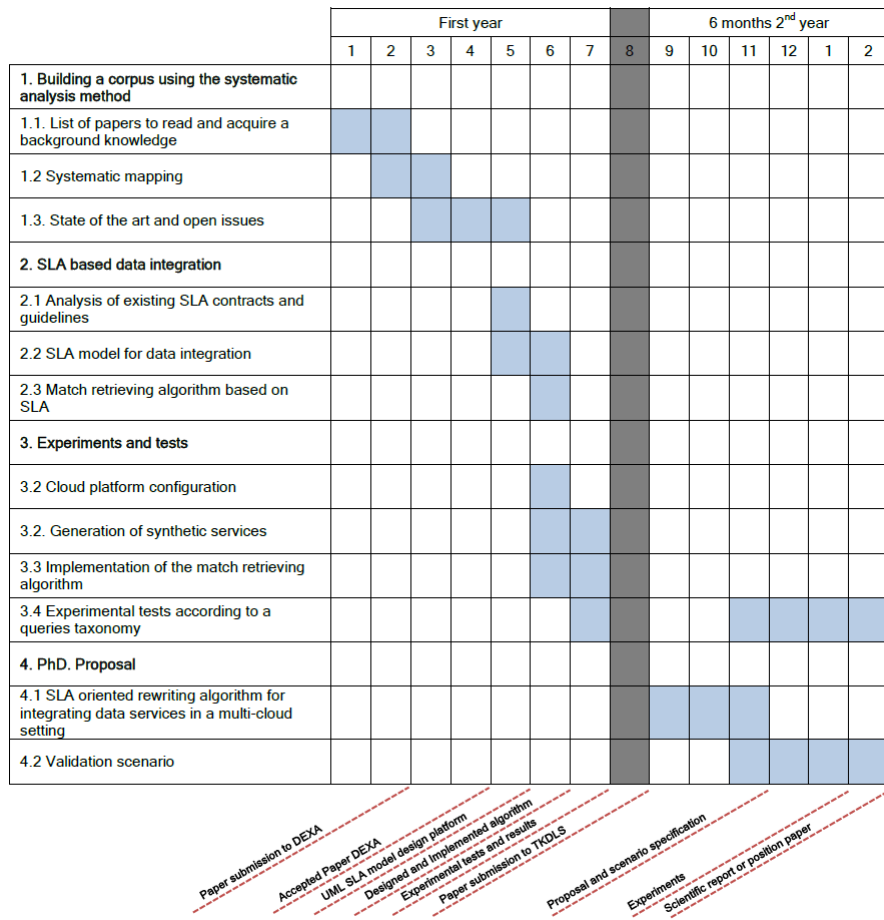
Figure 1: Calendar

Bibliography

[1] Mohammed Alhamad, Tharam Dillon, and Elizabeth Chang. Conceptual SLA framework for cloud computing. In *4th IEEE International Conference on Digital Ecosystems and Technologies*, pages 606–610. IEEE, April 2010.

[2] Ines Ayadi, Noemie Simoni, and Tatiana Aubonnet. SLA Approach for "Cloud as a Service". In *2013 IEEE Sixth International Conference on Cloud Computing*, pages 966–967. IEEE, June 2013.

[3] Karin Bernsmed, Martin Gilje Jaatun, Per Hakon Meland, and Astrid Undheim. Security slas for federated cloud services. *2012 Seventh International Conference on Availability, Reliability and Security*, 0:202–209, 2011.

[4] Schahram Dustdar, Reinhard Pichler, Vadim Savenkov, and Hong-Linh Truong. Quality-aware service-oriented data integration: Requirements, state of the art and open challenges. *SIGMOD Rec.*, 41(1):11–19, April 2012.

[5] Ghada ElSheikh, Mustafa Y. ElNainay, Saleh ElShehaby, and Mohamed S. Abougabal. SODIM: Service Oriented Data Integration based on MapReduce. *Alexandria Engineering Journal*, 52(3):313–318, September 2013.

[6] Hector Gonzalez, Alon Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, and Warren Shen. Google fusion tables: Data management, integration and collaboration in the cloud. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, pages 175–180, New York, NY, USA, 2010. ACM.

[7] Hector Gonzalez, Alon Y. Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, Warren Shen, and Jonathan Goldberg-Kidon. Google fusion tables: Web-centered data management and collaboration. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 1061–1066, New York, NY, USA, 2010. ACM.

[8] 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, New York, NY, USA, 2002. ACM.

[9] 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. COMPSAC 2007. 31st Annual International*, volume 2, pages 613–618, July 2007.