Daniel Aguiar da Silva Carvalho

2, rue Soeurs Bouvier • Lyon, France - 69005
E-mail: daniel.carvalho@univ-lyon3.fr

Current Position

PhD Student

I am a PhD student in Computer Science at the University Jean Moulin Lyon 3. Centre de Recherche Magellan's team.

Education

University Jean Moulin Lyon 3 [2014 – until now]

PhD in Computer Science at the University Jean Moulin Lyon 3.

Federal University from Rio Grande do Norte (UFRN) [2011 – 2013]

MSc. in Systems and Computation

Thesis entitled "Uma Máquina de Redução de Grafos para Serviços Web" (A Graph Reduction Machine for Web Services).

Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN) [2008 – 2011]

B. S. in Technology in Analysis and Systems Development.

Thesis entitled "MaRiSA-Design: Uma Ferramenta para Transformação do Modelo de Projeto Detalhado para Modelo de Linguagem de Programação" (MaRiSA-Design: A Tool for transformation from Design model to a Program Language model).

Projects

GT-CNC-2 [2013-2013]

The GT-CNC is a project funded by the Rede Nacional de Ensino e Pesquisa (RNP). In its first phase, the GT-CNC evaluated tools for building storage clouds and established a prototype based on one of the tested tools. The prototype is similar to cloud services like Dropbox which performs functions such as creating folders, sending files to cloud services, etc. For its second phase, the GT proposes to refine this prototype to conduct a pilot of the cloud storage service. This refinement includes the addition of authentication through the CAFe authentication and file access through the Web. In addition, the GT proposes to evaluate the adoption and integration of cloud with TERENA's Cloud Drive project.

Developer: As a team member I performed activities involving the following subjects: (i) Study and development of federated authentication; (ii) Extension of the Cloud Files API to support the federated authentication; and (iii) Integration of the Cloud Files API (which is implemented in Java) with the Cyberduck client (which is implemented using C#);

ePol [2010 - 2011]

ePol is a project to develop an electronic inquiry system to the Federal Police.

Developer: As a team member I performed activities involving the following subjects: (i) Development of corporative applications using JBOSS SEAM; (ii) Distributed version control using Mercurial; (iii) Development of corporative components with EJB (Enterprise Java Beans); (iv) Development of database components using JPA (Java Persistence API); (v) Development of web applications using JSF (Java Server Faces); (vi) Software testing with Selenium, jUnit and TestNG; (vii) Programming to Oracle databases; and (viii) Management of development projects with Redmine.

SMneticS [2010 - 2010]

The SMneticS is project to develop a web application for information disclosure through SMS messages.

Developer: the web application was developed using Java Server Faces and the PostgreSQL database.

Student Tech Club – RNet [2009 - 2009]

RNet is a study group formed by teachers and students that focus on developing application using the .Net technology.

Research leader: I was *group leader* during 1 year and I studied and taught lectures for beginners in programming using C# and Microsoft Visual Studio.

InterRed [2008 - 2010]

InterRed is a project that aims to create a system to provide, share, search and retrieval of a digital content aimed at vocational and technological education.

I was a *developer/designer* of flash applications and *digital content reviewer*.

Languages

Portuguese: mother language

English: read, written and spoken.

Publication

2015   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)

2014   Daniel A. S. Carvalho, José S. de Lima, Umberto S. Costa and Martin A. Musicante. A Graph Reduction Machine for Web Services. In preparation, 2014.

2013   Lucas Melo Silva, Felipe Leite da Silva, Daniel A. S. Carvalho, Thomas Diniz, Carlos Eduardo da Silva and Roberto Araújo. Estudo de Caso: Integração de Clientes da Nuvem Openstack Swift com uma Federação de Identidade. InWorkshop de Gestão de Identidades (WGID), XIII Simpósio Brasileiro de Segurança da Informação e Sistemas Computacionais (SBSEG 2013), 2013.

Teaching

Compilers. Federal University from Rio Grande do Norte (UFRN). Computer Science course. Brazil, 2011.

Theory of Computation. Federal University from Rio Grande do Norte (UFRN). Computer Science course. Brazil, 2013.

Practical Algorithms and Data Structured. Federal University from Rio Grande do Norte (UFRN). Computer Science course. Brazil, 2013.

Web Development. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Informatics' technical course. Brazil, 2013.

Structured and Object-Oriented Programming. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Informatics' technical course. Brazil, 2013.

Software Engineering. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Informatics' technical course. Brazil, 2013.

Applied Informatics. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Renewable Energy course. Brazil, 2013.

Fundamentals of Software Engineering. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Informatics' technical course. Brazil, 2014.

Structured and Object-Oriented Programming. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Informatics' technical course. Brazil, 2014.

Fundamentals of Logic and Algorithms. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Informatics' technical course. Brazil, 2014.

Informatics. Federal Institute of Education, Science and Technology from Rio Grande do Norte (IFRN). Electrical technical course. Brazil, 2014.

Programming Skills

.NET Framework *2 years experience*	Programming using Microsoft Visual Studio and SQL Server C# Programming Language Silverlight Asp.net
Java *6 years experience*	Programming using NetBeans IDE and Eclipse Java Programming Language Web development with JSP and JSF Corporative development with EJB Web service development Database access using JPA and JDBC Application development using JBOSS SEAM
Other experience	HTML, HTML5, Javascript, CSS C, C++, Ruby PostgreSQL, MySQL SVN and Git

-Centre de Recherche MAGELLAN-
IAE – Université Jean Moulin Lyon 3
6, Cours Albert Thomas – 69355 Lyon Cedex 08

Prof. Chirine GHEDIRA-GUEGAN   Lyon, 31 mai   2015
Mel. Chirine.ghedira-guegan@univ-lyon3.fr
Tel. (33) 04 78 78 76 41

Dr. Genoveva VARGAS SOLAR   Grenoble, 31 mai 2015
Mel. Genoveva.Vargas@imag.fr
Tel. (33) 04 76 82 72 48

Lettre d'appréciation des travaux de thèse de Mr Daniel Aguiar da Silva Carvalho

Mr Daniel Aguiar da Silva Carvalho est inscrit en première année de thèse, **financé par une allocation de recherche de la région Rhône Alpes ARC 2014 N°06** - Technologies de l'Information et de la communication et usages Informatiques Innovants.

Son travail de thèse s'intitule «Trusted SLA Guided Data Integration On Multi-Cloud Environments» et s'inscrit dans la thématique bien d'actualité relative à l'intégration de données et à la sécurité dans les nouveaux environnements à base de services à savoir le multi-cloud. Depuis le début de la thèse, Mr Daniel Aguiar da Silva Carvalho a effectué un travail sur les fondements du problème à traiter, et a adopté une approche systémique pour établir une cartographie des articles relatifs à la problématique de la thèse dans sa globalité mais aussi séparément au tour de l'intégration de données en terme d'environnement, de type de contribution et de qualité de données d'une part; et d'autre part d'usage des SLA et de leurs métriques. Mr Daniel Aguiar da Silva Carvalho a appliqué les 5 étapes de ladite approche connue sous le nom de « Systematic Mapping » pour ressortir et classer 114 travaux qui constituent le point d'entrée de son état de l'art qu'il a entamé, pour rédiger à court terme un survey. Par ailleurs, Mr Daniel Aguiar da Silva Carvalho a commencé l'installation de son environnement expérimental avec OpenStack permettant de déployer des infrastructures de cloud computing.

Les travaux effectués ont fait l'objet d'une présentation orale au séminaire du groupe SI du centre de recherche MAGELLAN (mars 2015) et d'une soumission : « Can Data Integration Quality be enhanced on Multi-cloud using SLA? » acceptée à la conférence internationale on Database and Expert Systems Applications, DEXA 2015.

Pour toutes ces raisons, nous sommes satisfaites de l'avancement et favorables à la continuité de ces travaux de thèse.

Genoveva Vargas Solar   **Chirine Ghedira-Guegan**

Thesis Advancement Report 2014-2015 (First Year)

Thesis title: Trusted-SLA Guided Data Integration 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[1]

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 representation of heterogeneous data by defining a schema (e.g., global and local as view approaches); (ii) tagging data with meta-data or by associating them to knowledge (e.g. semantic Web approaches); and (iii) 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 benefit 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 resources are not necessarily located in the same cloud. One cloud cannot be expected to provide the necessary resources to fulfill application requirements. With growing needs and requirements, 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, user quality and security requirements of the integration, and cloud heterogeneity in expressing and enforcing the corresponding clauses [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 (i) on the lifecycle of a security SLA on hybrid clouds [3]; (ii) on SLA models for addressing management capabilities as a service, Pcloud services, performed through agreed and negotiated in contracts (elasticity, high availability, scalability and on demand provisioning) [2]; and (iii) on 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].

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 all 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: 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; and that (iv) cloud services and data services are listed in a registry.

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.

3 Objectives

---

[1]You can find the references at https://www.dropbox.com/s/dyg08a622ucv6xl/references.pdf?dl=0

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 linked to data quality, data security and to cloud resources; (2) Propose an unified formalism to represent them; (3) 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 (4) 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 consisting in retrieving papers from scientific databases, filtering them according to inclusion and exclusion criteria and research interests expressed in research questions. A classification schema was proposed, consisting in facets and dimensions. The abstracts of the final papers collection were read to classify each paper within the scheme. We identified the trends and open issues 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 collection of 114 papers and analytics results. 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 using Open Stack to implement and evaluate our match-retrieving algorithm [2].

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 (http://fr-br-school.imag.fr).

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

---

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