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.

Thesis title: Trusted SLA Guided Data Integration on Multi-Cloud Environments.

Supervisor: Chirine Ghedira-Guegan.

Co-supervisors: Nadia Benani and Genoveva Vargas-Solar.

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.

French: basics.

Publication

2016 D. A. S. Carvalho, P. A. Souza Neto, G. Vargas-Solar, N. Bennani, C. Ghedira. Towards Quality Guided Data Integration on Multi-Cloud Settings, Service Oriented Computing - 14th International Conference, ICSOC 2016, PhD Symposium, Banff, AB, Canada, October 10-13, 2016.

2016 D. A. S. Carvalho, P. A. Souza Neto, G. Vargas-Solar, N. Bennani, C. Ghedira. Rhone: a quality-based query rewriting algorithm for data integration, Short paper, 20th East-European Conference on Advances in Databases and Information Systems, ADBIS 2016, pp. 80-87.

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.

2015 M. A. Macedo, D. A. S. Carvalho, M. A. Musicante, A. Pardo and U. S. Costa. An abstract machine for integrating heterogeneous web applications, 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), Marrakech, 2015, pp. 1-8.

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.

Events

2015 — 1st French Brazilian Spring School on Smart cities and Big data at University of Grenoble Alpes, Grenoble, France.

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

-Laboratoire LIRIS CNRS (UMR 5205)-
IAE – Université Jean Moulin Lyon 3
6, Cours Albert Thomas – 69355 Lyon Cedex 08

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

Lyon, 20 mai 2017

Dr. Genoveva VARGAS SOLAR
Mel. Genoveva.Vargas@imag.fr
Tel. (33) 04 76 82 72 48

Grenoble, 20 mai 2017

Avis sur le rapport d'activité de Mr Daniel Aguiar da Silva Carvalho

Mr Daniel Aguiar da Silva Carvalho est inscrit en troisième 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 décrite dans le rapport joint sur l'intégration de données dans un contexte de confidentialité dans les nouveaux environnements à base de services à savoir le multi-cloud. Durant cette année, le candidat a consacré son travail à la finalisation de l'intégration de clauses de préférences et des clauses SLA et sa mise en œuvre dans un algorithme de réécriture. A cet effet, il a proposé une extension à l'algorithme Rhône qui lui a permis de mieux comprendre/appréhender la notion d'intégration en composant des services de données qui doivent satisfaire des clauses représentant les besoins d'un consommateur de données, et à montrer l'efficacité et le gain de temps.

Par ailleurs, Mr Daniel Aguiar da Silva Carvalho s'est ensuite attelé au challenge imposé par le contexte multi-cloud et relatif aux schémas des différents types de SLAs pouvant exister dans un tel environnement, amenant une difficulté supplémentaire à ladite intégration du fait des hétérogénéités possibles, des manques éventuelles de clauses, et de la génération de SLA de dérivation et d'intégration pour améliorer la réutilisabilité. Cette phase constitue l'originalité du travail de thèse et pour laquelle Mr Carvalho a proposé un SLA, un accord entre un service de données et un fournisseur de cloud, et un SLA de service qui est un nouveau type d'accord défini par les services de données exposant les propriétés des données qu'ils fournissent.

De plus, afin d'améliorer les performances et réduire le temps d'intégration, le candidat a proposé une définition et formalisation d'une taxonomie de requêtes ainsi qu'une méthode de réutilisation des requêtes précédentes.

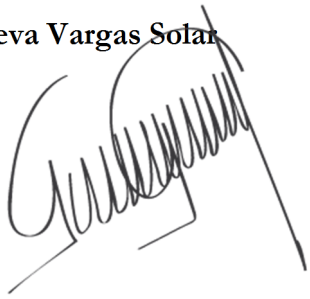
Par ailleurs, le candidat a bien effectué des présentations orales aux séminaires de groupe et d'équipe de recherche durant cette année, ainsi qu'aux journées scientifiques de l'ARC6. Il est en

cours de rédaction et soumission d'un article portant sur la réutilisation des requêtes dans le Multi-Cloud.

En conclusion sur cette troisième année, Mr Daniel Aguiar da Silva Carvalho a montré son aptitude pour une recherche applicative et a fait preuve de motivations pour apporter un équilibre entre résultats théoriques et techniques (par exemple l'implémentation de l'algorithme versus l'énoncé formel du problème et d'une approche). Nous espérons ainsi que Mr Daniel Aguiar da Silva Carvalho pourra l'accomplir selon le planning proposé dans son rapport. Vu son état d'avancement, nous prévoyons un dépôt de la thèse de Mr Daniel Aguiar da Silva Carvalho d'ici décembre 2017.

Mr Daniel Aguiar da Silva Carvalho bénéficiera d'un financement sur un contrat d'ingénieur d'études au Brésil pour un montant de 1200€.

Genoveva Vargas Solar



Chirine Ghedira-Guegan

Thesis Advancement Report 2016-2017 (Third 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

User-guided data integration implies consuming data by matching and composing different data services and integrating the results while respecting data consumers quality requirements. These requirements could be defined in service level agreement (SLA) contracts established between data consumers and data providers. The SLA defines what a data consumer can expect as system behavior, but also the properties of the data such as its provenance, veracity, freshness, whether the consumer accepts to pay for data, and how much is he/she ready to pay for the resources necessary for integrating his/her expected result.

Several authors have introduced algorithms for data integration in service-oriented architectures (such as [Barhamgi 2010, Costa 2013, Benouaret 2011, Ba 2014]). They have focused on the query-rewriting problem, which in this context includes matching, selecting and composing services according to some data requirements and constraints. The problem of composing services has been proven to be an NP-hard problem, which implies a performance problem while searching for service combinations. To tackle it, heuristics for computing and identifying the best services and compositions based on quality aspects have been proposed (such as [Cardoso 2004, Berbner 2006, Menascé 2008, Sasikaladevi 2014]). However, current work focus on services' properties neglecting data properties and the new constraints imposed by the service-oriented context (for example, a given data service could be out of resources according to what he agreed on SLA with his service provider).

The service-oriented context brings challenges to query rewriting in data integration solutions. Instead of taking into consideration only the user query and his/her requirements with respect to services' properties (such as percentage of availability and response time), the integration process must consider in addition the new constraints imposed by the context:

- User requirements may concern not only data services' properties but also quality requirements of the data, which is being provided (such as freshness, cost, provenance, data type, veracity among others).
- Data provision is constrained to the available computing resources agreed between data services and service providers on service level agreement (SLA) contracts. For example, a data service could have agreed to perform a limited quantity of requests per day.
- The data integration process requires a high level of computing resources while searching for services and producing compositions. The huge amount of data and data services in the service context increases even more the complexity of the solutions.

Concerning the aforementioned, current data integration solutions introduce a multi-dimensional matching problem that should take into account:

- Matching and selecting data services according to the data consumer requirements (concerning service and data's properties) with respect (i) to data services quality measures defined in SLA contracts; and (ii) to data services' available resources according to different SLAs that they have agreed with different service providers.
- Delivering results with respect to the data consumer requirements depending on the context which he/she consumes the data (for example, using a mobile phone)

Thus, this thesis addresses data integration on service-oriented environments. The aim is to propose a data integration approach targeting (multi-)cloud context in which data is delivered according to data consumers' expectations and financial constraints. The multi-cloud introduces a new vision to the problem considering that data services could deploy services in different cloud providers, under different quality conditions (with respect to performance and data properties) which are agreed in several SLA contracts under different pricing conditions.

Furthermore, the explosion in the number of services increases the complexity of service selection and combination, and the decision of which rewriting will be done according to the different quality aspects and cost. To achieve this, this thesis aims to fulfill these three main contributions:

- A new query rewriting algorithm which takes into account user requirements regarding data properties and service properties according to the deployment context (see appendix B);
- Enhancing the algorithm by proposing an heuristic-based approach to avoid the combinatorial problem while searching and composing services (Future work); and
- Reducing the rewriting frequency by building a framework that reuses previous data integration based on a query history (work in progress).

2. Synthesis of the Research Activities

During the third year, we have organized our activities as follows:

Data integration meta-model and SLA schemas: to describe our context and to illustrate the approach, we have designed a meta-model for data integration. Current SLA schemas are focused on service properties aspects such as availability and response time. Thus, to better fit on the data integration requirements, we have proposed cloud SLA that is an agreement between a data service and a cloud provider, and a service SLA which is a new kind of agreement defined by data services exposing the properties of the data they provide. The results of this part of our work have been accepted in the ICSOC PhD Symposium 2016.

A method for service and composition selection: Current data integration implies dealing with a huge number of data services. Consequently, query rewriting activities become more expensive in terms of computing resources and generation time. In this scenario, it is mandatory to develop a method to identify the best services to produce the best compositions that can achieve the user satisfaction. Thus, we have started working in an heuristic to rank data services and compositions based on SLA measures concerning service properties (percentage of availability, response time, throughput and others) and data properties (data type, freshness, veracity, provenance and others).

Definition and formalization of a taxonomy of queries: We have defined and formalized a set of possible relations between queries, which differ in terms of abstract services, service properties and data properties.

A method for reusing queries: It is well known that query rewriting is expensive. Thus, based on the proposed query taxonomy, we have designed and formalized a reusability approach, which allows reusing data services and compositions from previous integration in order to profit from them.

Query history data model and implementation: The reusability approach is based on a history of previous integrated queries. While defining the query taxonomy and reusability issues, we have identified the key information that should be part of the history allowing a satisfactory reusability process. A query history data model, which includes queries, abstract services, data services and compositions, was designed using the collected information. Further, the Rhone algorithm was adapted to be in accordance with the model.

3. Perspectives

Our perspective is to tackle what is remaining in the contributions while writing the final thesis document. Our intended calendar presented below:

Activities:	**Feb - April**			**May - July**			**Aug - Dec**				
Formalization of the query taxonomy and reusability functions											
Implementing reusability functions and including in the *Rhone*											
Building a proof of concept to the approach											
Writing a scientific paper											
Optimizing the first version of the approach											
Heuristics for optimizing the service selection and composition											
Writing the final thesis document											
Thesis Defense											

Rapport du comité de suivi individuel – Ecole Doctorale InfoMaths

Nom & prénom(s) du (de la) doctorant(e) : Daniel Aguiar da Silva CARVALHO

Laboratoire : Magellan **Etablissement :** Lyon 3

Directrice(s)/Directeur(s) de thèse : C. GHEDIRA - GUEGAN
G. VARGAS-SOLAR N. BENNANI

Titre du projet de thèse : Trusted SLA Guided Data Integration on Multi-Cloud Environments

Année d'inscription : D.. 2014 **Date du début de la thèse :** 1er oct. 2014.

Co-tutelle ☐ **CIFRE** ☐

NB : Pour les thèses en partenariat industriel ou international, le comité veillera à l'équilibre entre les partenaires (communication, cohérence, présence sur chaque site, répartition du temps).

Composition du comité :

Personnalité(s) extérieure(s) : M. Ferrari

Personnalité de l'ED InfoMaths : E. Duchêne

Date du comité : 28.03.17

CRITERES	AVIS -	AVIS +	COMMENTAIRES
Avancement des travaux de recherche	☐	☒	
Publications	☐	☒	1 travail à finir
Présentations orales	☐	☒	
Formations doctorales	☐	☒	
Intégration au sein du laboratoire	☐	☒	
Conditions matérielles	☒	☐	manque de matériel au laboratoire (pas de serveur)
Interactions entre le(la) doctorant(e) et l'équipe encadrante (fréquence des réunions, communication, réactivité)	☐	☒	
Activités complémentaires	☐	☒	pas d'enseignement à cause de la langue.

- : point faible + : point fort

		Acceptées	Soumises
Publications	Revues Internationales : Revues Nationales : Conférences Internationales : Conférences Nationales : Autres :	/ 4 /	
Présentations orales	Séminaires : / Conférences : 2 Posters : 2 (ARC) Autres : 5 (local)		
Formations doctorales	Nombre heures FSC : 117 + 39. OK. Nombre heures FIP : Au Brésil, en attente de validation.		
Activités complémentaires	Enseignement (contrat, vacations) : Mandat électif : Autres : /		

Progression du projet de thèse et appréciation globale

Bonne progression, soutenance prévue en septembre.
De bonnes publications.

Recommandations pour l'année suivante

1 publication journal serait bienvenue.
Propose plus de comparaisons avec les méthodes existantes.

Avis pour la réinscription en D..

Avis favorable ☒

Avis favorable sous réserve ☐

Avis défavorable ☐