Daniel Aguiar da Silva Carvalho

124, cours Gambetta • Lyon, France - 69007
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, 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 (to appear).

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 — Daniel A. S. Carvalho, José S. de Lima, Umberto S. Costa and Martin A. Musicante. A Graph Reduction Machine for Web Services. In preparation, 2015.

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, 1 juin 16

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

Grenoble, 1 juin 16

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

Mr Daniel Aguiar da Silva Carvalho est inscrit en seconde 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 l'étude de faisabilité 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 proposé par l'équipe de recherche de l'Université Rio Grande do Norte, Brésil, 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.

Cette partie étant pratiquement finalisée, Mr Daniel Aguiar da Silva Carvalho devrait s'atteler maintenant 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. Par ailleurs, le candidat a bien effectué 3 présentations orales aux séminaires de groupe et d'équipe de recherche durant le premier semestre, a soumis un article qui a été accepté à la conférence internationale ADBIS 2016 et est en cours de soumission d'un papier au Phd symposium de la conférence internationale ICSOC 2016.

En conclusion sur cette seconde année, Mr Daniel Aguiar da Silva Carvalho a plus montré son aptitude pour une recherche applicative et une tendance à aborder les problèmes de façon empirique sans se référer à l'étude de l'existant élaborée l'année précédente. Ce qui explique un manque de moyens pour aborder entièrement le problème scientifique de la thèse, et a

engendré un déséquilibre entre les résultats purement théoriques et techniques (par exemple l'implémentation de l'algorithme versus l'énoncé formel du problème et d'une approche). Compte tenu de ces constations, et malgré le manque d'attitude proactive lors des discussions scientifiques avec les encadrantes, dû probablement à sa culture, nous avons et continuons à encourager Mr Daniel Aguiar da Silva Carvalho à avoir une stratégie et une méthodologie de travail plus mature et professionnelle de façon à pouvoir aborder les défis susmentionnés pour compléter et finaliser son travail de thèse. Nous espérons ainsi que Mr Daniel Aguiar da Silva Carvalho pourra l'accomplir selon le planning proposé dans son rapport.

Genoveva Vargas Solar

Chirine Ghedira-Guegan

Thesis Advancement Report 2015 – 2016 (Second 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

Cloud architectures have been widely used for hosting different types of resources and applications. Current data integration solutions imply consuming data from different data services, integrating the results, and delivering them to the consumer in a homogeneous way. Nowadays, these services could be deployed in different clouds (multi-cloud) profiting from the elasticity, distributed processing and *pay-per-use* model imposed by this scenario. However, the multi-cloud context adds another complexity to the integration process, as it imposes and introduces new constraints and characteristics associated to data consumer, data producers, the cloud infrastructure and the data itself. Such constraints and characteristics are related to data security, data quality, processing and memory limits, storage, budget, among others.

Thus, given a consumer query, the integration process in this context should match and take into consideration all the constraints and characteristics of each entity (data consumer, data producers, the infrastructure and the data) while delivering the results to the consumer. To achieve this, we believe that service level agreements (SLA) can be used to unify all constraints and requirements information to serve as a meaning for the integration. However, the current SLA models are not able to cover all the integration aspects and the new issues in this process. Due to this reason, the first challenge is to propose a new kind of SLA (called Integration SLA) that matches the user's integration preferences (including constraints and characteristics) with the SLA's provided by cloud services, given a specific user cloud subscription. In this context, the matching process can lead to (i) an exhaustive search in the chain of SLAs; (ii) deal with SLA incompatibilities; and (iii) deal with heterogeneous SLA specifications (different schemata, different measures semantics and granularities).

Another challenge is to guide data integration taking into consideration the integrated SLA. Here, the data integration process includes (i) looking up services that can be used as data providers, and for services required to process retrieved data and build an integrated result; (ii) performing data retrieval, processing and integration and (iii) deliver results to the user considering his/her requirements (for example, concerning quality, freshness, context and resources consumption). The integrated SLA guides services selection and filtering considering all the constraints and characteristics of each actor. In addition, it helps to control the amount of data to be retrieve and processed according to the user's consumption rights.

The objective of this PhD is to propose data integration strategies adapted to the economic model of the cloud taking into account all the constraints and characteristics imposed by the multi-cloud context.

2. Synthesis of the Research Activities

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

1. **Development of our data integration approach**. We have been working on our data integration approach, which is briefly described as follows. Given a query and a set of user requirements associated to it, the query execution process is divided in three phases. The first phase creates a SLA for the user request. It consists in looking for a (stored) integration SLA for a similar request. If a similar SLA is found, it is reused and the request is forwarded to the query evaluation phase. Otherwise, a new SLA to the integration (called integration SLA) is produced. The query is expressed as a service composition with associated user preferences and constraints. In the second phase, service composition, the query is rewritten in terms of different services considering the user preferences and the SLAs of each service involved in the composition. The rewriting result is stored for further uses. Finally, in the query evaluation phase, the query is optimized in terms of user preferences and SLAs concerning the consumed resources and the economic cost of the query. Once optimized, the query processed in the execution engine.
2. **Extension of the query-rewriting algorithm**. In collaboration with our colleagues in Brazil (authors in [1][1]), we have worked on an adapted version of [1] to our data integration solution extending their data structure to map services to the query, and adding the concepts of user preferences to the query and quality measures to the services. In addition, we have developed and formalized the *Rhone* service-based query-rewriting algorithm. The algorithm extension proposes two original aspects: (i) the user can express his/her quality preferences and associated them

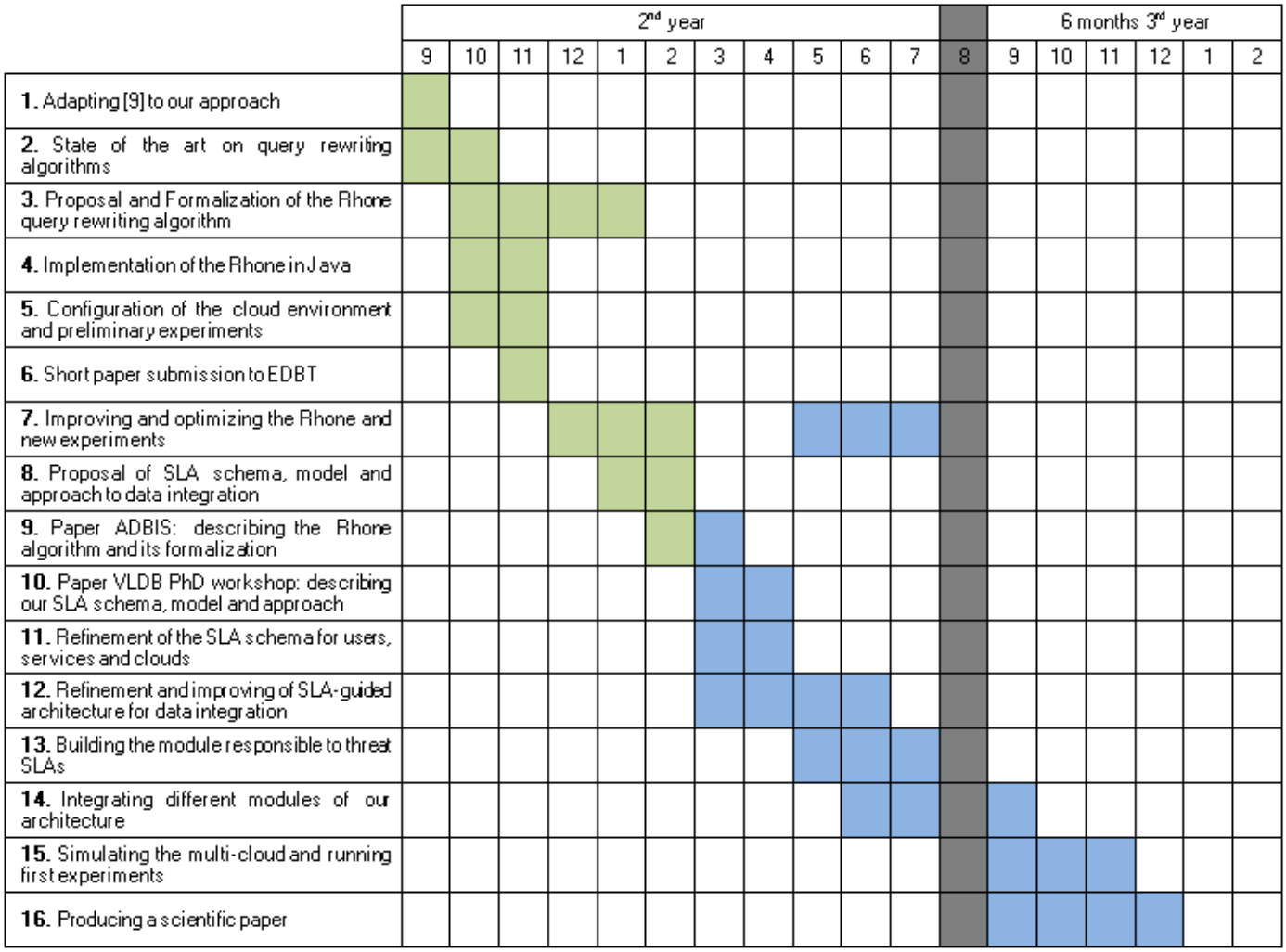
[1] Ba, C., Costa, U., H. Ferrari, M., Ferre, R., A. Musicante, M., Peralta, V., Robert, S.: **Preference-driven refinement of service compositions**. In: Int. Conf. on Cloud Computing and Services Science. Proceedings of CLOSER 2014.

to his query; and (ii) service's quality aspects defined in SLAs guide the service selection and the whole rewriting process taking into consideration that services and rewritings should meet the user requirements, and the different cases of incompatibilities of SLAs, uncompleted SLA and the integration SLA. Preliminary experiments were produced to evaluate the *Rhone*.

3. **SLA Model**. We have been working on the SLA model to data integration. The model considers different entities (such as data producers, data consumers, the infrastructure and the data). Each concept has its constraints and characteristics: (i) data producers describe the type of data they produce, production rate and time, location, cost, and they have an associated SLA defining access policies and resources limits over the infrastructure; (ii) data consumers have could subscriptions defining access policies, resources limits, but they also describe their requirements in terms of data quality, budget and privacy aspects; (iii) the infrastructure is characterized in terms of processing limit, storage, memory and access policies; and (iv) the data itself can be tagged considering its type, security issues, and quality aspects (such as provenance, freshness, degree of rawness, and veracity). Each concept and its constraints and characteristics are taken into account during the integration process. In the next step, we are going to analyze how the constraints and characteristics can be represented on SLA in order to be mapped and matched during the integration.
4. **Publications**.
    - 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 (to appear).
    - In addition, we are working on another paper to be submitted to ICSOC PhD Symposium (8 June).
5. **Presentations**. In order to have an external feedback, we have presented our work during the monthly group meeting in the SOC-Team research group from Lyon 1.

3. Perspectives

Currently, SLA incompatibilities are not taken in account. We are working on this issue, enriching our model and approach. We have also to work and study heuristics to be applied in our algorithm in order to reduce the composition search space making in this sense the integration more efficient. Moreover, it is necessary to analyze how the query execution plan should be parallelized to let the execution more efficient in a multi-cloud environment. Finally, evaluate and validate the entire quality-based data integration approach in a multi-cloud configuration. The calendar below describes our intended activities.



ENS DE LYON — Université Claude Bernard Lyon 1 — CENTRALE LYON — UNIVERSITE LUMIERE LYON 2 — INSA LYON

Ecole Doctorale Informatique et Mathématiques

InfoMaths ED 512
Directrice : Sylvie CALABRETTO

Tèl. : 04 72 43 80 46
Fax : 04 72 44 80 53
E-mail :infomaths@infomaths.univ-lyon1.fr
Internet : http://infomaths.univ-lyon1.fr

A qui de droit

Villeurbanne, le 1er septembre 2014

ATTESTATION

Je soussignée, Sylvie CALABRETTO, Directrice de l'Ecole Doctorale Informatique et Mathématiques (ED InfoMaths), atteste que Monsieur **Daniel AGUIAR DA SILVA CARVALHO né le 16 septembre 1988 au BRESIL** est autorisé à s'inscrire en tant que doctorant en Informatique pour l'année 2014-2015 à l'ED InfoMaths.

Le candidat bénéficiera d'un financement (Allocation de recherche de la région Rhône Alpes ARC 2014 N°06) d'un montant de 2537€ /mois sur 3 ans.

La directrice de thèse sera Chirine GHEDIRA GUEGAN.

.

Sylvie CALABRETTO,
Directrice InfoMaths

InfoMaths Informatique et Mathématiques
LYON I
LYON II
INSA LYON
EC LYON
ENS LYON
ED 512