

1 Equipe HADAS - Axe Traitement de Données et de Connaissances à Grande Echelle

1.1 Scientific Presentation

Research group:

The Hadas group was formed in october 2005 as a new team for the LIG laboratory. It actually follows the STORM team, directed by M. Adiba until 1996. Over the years, the group proposed an evolution of the scientific vision of a database management systems as a semantic-based infrastructure for managing ubiquitous and heterogeneous data services. The team currently includes 9 research persons (2 full professors, 5 Associate Professor and two CNRS Research Scientist) and a research engineer and 15 PHD students and post-docs. Durant this period we hired 3 persons.

The research activity on services has been reinforced in the group with the arrival of Noha Ibrahim, associate professor who joined the team in October 2010. Then, the arrival of Sihem Amer-Yahia Research (Dec 2011) and Vincent Leroy Associate Professor UJF / Research scientist (Sept 2012) brought new areas of research in the group. Etienne Dublé Research Engineer (Mai 2011) helps us in developing and finalizing some of our prototypes. He also brought expertises in new kind of sensor networks and of file systems.

Research description and :

The HADAS group has contributed in the following areas: relational data models, snapshots and their semantics, active and temporal databases and object-oriented database systems. The advent of the web and middleware infrastructures in the early 1990s has profoundly changed the nature of research in databases. Our research is related to the changes in devices and softwares:

- Memories and disks with more capacity and faster access
- faster processors and networks
- Algorithmic advances, eg , parallel computing
- cloud computing (virtualization , elasticity, pay- as-you -go ...)

More recently big data management and analysis gives other more challenging prospective. These changes have reduced the cost of creating, capturing, managing and storing information of a sixth of what it was in 2005. This has allowed a scale change in size of data, distribution of data, number of connected devices, and number of users.

Considering this globalization of data, knowledge and computing resources and because of the quality and intelligence of the expected data management functions, we have been confronted (and still confronted) with a change in the field of databases.

To face this challenges we decide to revisit database systems and consider them not anymore as centralized data storage systems but data management services largely distributed and deployed over different types of large scale systems (grids, peer-to-peer networks, sensor networks, ambient and ubiquitous environments). Semantics is at the heart of this approach as it is used at all levels

of the process of designing or composing data services for handling autonomy, dynamic behavior and heterogeneity of both users and data sources.

The activities of the group during these past years have been centered on the following themes:

- Accessing data in large-scale systems: a first aspect concerns query optimization in distributed and dynamic systems ; a second aspect deals with mining large amounts of data to extract patterns of interest.
- Composing data services in a dynamic way: we investigate models, algorithms and tools for coordinating services with non functional properties (contracts) and for providing access to heterogeneous data coming from services
- Reasoning on data semantics: we investigate different models and algorithms for querying data (or resources) through possibly heterogeneous and distributed ontologies.

We have participated to the Optimacs and Ubiquist ANR projects that bridge the gap between data management and (web) services into a data management platform and between networks and data management.

We have been also involved in other ANR projects on context management for software adaptation, handling uncertainty and trust in peer-to-peer data management system. We are involved in the Datalyse project whose objective is to build a smart warehouse demonstrator for the collection, certification, integration, categorization, security, enrichment and sharing of heterogeneous Big Data of type "Big Data User" (UBD) or from machines, of type "Big Data Monitoring" (MBD). We also collaborate with industry, especially with ST Microelectronics on novel data mining algorithms able to fully exploit the parallelism of multicore machines.

Results of our research have direct impact on applications dealing with huge amounts of data and resources largely distributed in pervasive environments, such as data spaces, smartgrids and smart buildings, hardware and software observation, and the semantic Web.

1.2 Scientific and Technological Results

1.2.1 Accessing data in large-scale systems

Accessing data concerns several dimensions of large scale systems: number of resources, data volume and data complexity. It basically means using declarative queries that are optimized based on system characteristics. Data mining is another way to query large quantities of data, by extracting interesting patterns from them. Such patterns provide meaningful abstractions of raw data, which are thus less numerous and more appropriate for data analysis. Globally, the difficulty for evaluating queries efficiently on nowadays applications motivates this work to revisit traditional query optimization techniques. The following presents these two aspects of accessing data in the large. It also focuses on works done on querying the social web.

1- CBR query optimization

Our research contributes to the development of new distributed query optimization techniques. It relies on the adaptation of machine learning, more precisely Case-Based Reasoning(CBR), and pseudo random search space exploration (also exploiting the case base) to produce efficient query execution plans according to application specific optimization objectives expressed over resource consumption (e.g. time, energy, number of messages). The query plan generation considered multiple optimization objectives customizable to application requirements (QoS based Hybrid Query optimization)

These research led to the following original contributions:

- A query optimization approach that use cases generated from the evaluation of similar past queries. A query case comprises: (i) the query (the problem), (ii) the query plan (the solution) and (iii) the measures of computational resources consumed during the query plan execution (the evaluation of the solution).
- A query plan generation process [1] that uses classical query optimization heuristics and makes decisions randomly when information on data is not available (e.g. for ordering joins, selecting algorithms or choosing message exchange protocols). This process also exploits the CBR principle for generating plans for subqueries, thus accelerating the learning of new cases.
- A Simulation Platform [2] allowing to experiment distributed query optimization and rule-based programs over a set of distributed data-enabled devices hosting virtual machines(VM). A VM integrates a query optimization engine [3] implementing the above techniques.

Contracts: ANR projects: UBIQUEST, OPTIMACS (coordinators); AGIR project Wait; FP7 EU project Cases(coordinator).

Key references:

- [1] Lourdes Martinez, Christine Collet, Christophe Bobineau, Etienne Dublé. The QOL approach for optimizing distributed queries without complete knowledge. IDEAS, 91-99, 2012
- [2] Ahmad Ahmad-Kassem, Christophe Bobineau, Christine Collet, Etienne Dublé, Stéphane Grumbach, Fuda Ma, Lourdes Martinez, Stéphane Ubéda. UBIQUEST, for rapid prototyping of networking applications. IDEAS, 187-192, 2012
- [3] Lourdes Martinez, Christine Collet, Christophe Bobineau and Etienne Dublé. CoBRa for optimizing global queries. BDA, 2013
- [4] Carlos-Manual Lopez-Enriquez, Genoveva Vargas-Solar, José-Luis Zechinelli-Martini, Christine Collet. Hybrid query generation, LANMR, 117-128, 2012

2- Data mining

Data mining is the automatic extraction of unknown and potentially interesting information from large quantities of data. One of the major fields of data

mining consists in discovering patterns occurring frequently (i.e. more than a given threshold) in data. The group works on pattern mining in complex data such as sequences, trees or graphs, which are found in many applications in chemistry (e.g. graphs representing molecules) or in bioinformatics (e.g. gene regulation networks).

This research focused on improving frequent pattern mining algorithms, both to make them more scalable and to apply them to real data analysis contexts. Main results are:

- From the scalability point of view, we acquired a strong expertise on exploiting multicore processors for pattern mining [2, 5]. The proposed algorithms are also based on the notion of closed patterns [2, 4, 5], reducing the output size (hence the computation time) without loss of information.
- These works culminated with the proposition of ParaMiner [2], the first parallel and generic algorithm for mining closed patterns.
- From an application point of view, works have been done on the analysis of execution traces, in collaboration with STMicroelectronics. They have improved the way to discover periodic behaviors and their disruption in traces [4], to rewrite a trace with a few significant sequences of events [3], and to automatically discover hotspots of memory contention in a parallel code [1].

Contracts: FUI SoCTrace, leader for UJF partner

Key references:

- [1] Sofiane Lagraa, Alexandre Termier, Frédéric Pétrot: Data mining MPSoC simulation traces to identify concurrent memory access patterns. DATE 2013: 755-760
- [2] Benjamin Négrevergne, Alexandre Termier, Marie-Christine Rousset, Jean-Francois Méhaut: ParaMiner: a generic pattern mining algorithm for multi-core architectures, Data Mining and Knowledge Discovery, 2013
- [3] Christiane Kamdem Kengne, Leon Constantin Fopa, Alexandre Termier, Noha Ibrahim, Marie-Christine Rousset, Takashi Washio, Miguel Santana: Efficiently rewriting large multimedia application execution traces with few event sequences. KDD 2013: pp 1348-1356
- [4] Patricia Lopez-Cueva, Aurélie Bertaux, Alexandre Termier, Jean-Francois Méhaut, Miguel Santana: Debugging Embedded Multimedia Application Traces through Periodic Pattern Mining, EMSOFT 2012 pp 13-22
- [5] Trong Dinh Thac Do, Anne Laurent, Alexandre Termier: PGLCM: Efficient Parallel Mining of Closed Frequent Gradual Itemsets, ICDM 2010 pp 138-147

3- Social Web Data access

Research has been done on new exploration problems to find useful user groups in collaborative rating datasets [1,2,4] and useful information in online news [3,5]. Our formulation of exploration as an optimization problem where various dimensions such as similarity, diversity, and coverage are optimized, leads

to reductions from well-known problems and adaptations well-established algorithms. Large-scale user studies have been conducted to verify the effectiveness of our findings. The current research direction is to blend efficient mining with exploration and to develop an evaluation methodology for large-scale information exploration.

Social Web Data access also concerns optimization. Data is stored within data centers, which constitute distributed systems. It is therefore important to optimize communications between the machines to avoid saturating the network equipment. A first work considered the problem of data routing between users of social networks. The key idea was to identify hubs that aggregate data from several sources and reduce the number of messages exchanged [6]. A second work considers the problem of data placement in hierarchical network structures. A reactive algorithm monitors data access patterns to identify locations in which new replicas of data should be deployed to reduce routers saturation [7].

Contracts: Datalyse: Big Data Models and Algorithms, Investissement d’Avenir 2013-2016, ; ALICIA: ANR 2014-2017. AGIR 2013.

Key references:

- [1] Behrooz Omidvar Tehrani, Sihem Amer-Yahia, Alexandre Termier, Aurélie Bertaux, Éric Gaussier, Marie-Christine Rousset: Towards a Framework for Semantic Exploration of Frequent Patterns. IMMoA 2013: 7-14 (workshops)
- [2] Mikalai Tsytarau, Sihem Amer-Yahia, Themis Palpanas: Efficient sentiment correlation for large-scale demographics. SIGMOD Conference 2013: 253-264
- [3] Sofiane Abbar, Sihem Amer-Yahia, Piotr Indyk, Sepideh Mahabadi: Real-time recommendation of diverse related articles. WWW 2013: 1-12
- [4]ahashweta Das, Saravanan Thirumuruganathan, Sihem Amer-Yahia, Gautam Das, Cong Yu: Who Tags What? An Analysis Framework. PVLDB (11): 1567-1578 (2012)
- [5] Demo: Sihem Amer-Yahia, Samreen Anjum, Amira Ghenai, Aysha Siddique, Sofiane Abbar, Sam Madden, Adam Marcus, Mohammed El-Haddad: MAQSA: a system for social analytics on news. SIGMOD Conference 2012: 653-656
- [6] Aristides Gionis, Flavio P. Junqueira, Vincent Leroy, Marco Serafini and Ingmar Weber: Piggybacking on social networks. In Proceedings of the 39th International Conference on Very Large Databases (VLDB), pages 409-420, 2013
- [7] Xiao Bai, Arnaud Jégou, Flavio P. Junqueira and Vincent Leroy: yna-SoRe: Efficient In-Memory Store for Social Applications. In Proceedings of the 14th International Middleware Conference (Middleware) pages 425-444, 2013

1.2.2 Composing data services on the fly

Composing services exported by different organisations is a key issue when building large scale and data-intensive systems. Composition requires to take into

account the characteristics of eco-systems (e.g., memory and computing, and network capabilities). The composition process uses this knowledge or semantics to dynamically discover and coordinate (ubiquitous) services, and then to adapt the coordination process depending on the availability and change of services. Another important challenge is to consider non functional aspects and QoS (quality of service) criteria such as availability, reliability, and temporal constraints that are crucial when composing data services in a dynamic way. We also have to consider an unlimited set of resources usable via a "pay as U go model", energy consumption or services reputation and provenance models.

The research in this theme contributes to the construction of service based data management systems as service coordinations. Once data management is delivered as service coordinations, it can have associated non-functional properties: exception handling and recovery. We proposed methodologies, algorithms, languages and tools for designing and executing service coordinations respecting SLA contracts.

We first applied our approach for the management of queries as coordinations of services, including data services and computing. This leads to the following results:

- an Hybrid query model for expressing queries as data service coordinations based on workflows. It uses the abstract state machines (ASM) formalism.
- a query language HSQL (Hybrid Services Query Languages) associated to the hybrid query model and the language MQLiST (Mashup Query Spatio Temporal Language) for integrating hybrid query results in a mashup. Both are extension of SQL.
- an algorithm BP GYO for generating the query workflow that implements a query expressed in HSQL.
- an hybrid query evaluation engine HYPATIA.

Another result of our research is linked with the QoS of service coordinations using active policies linked with activities of workflow .

Contracts: OPTIMACS (2009-2012) Agence Nationale de la Recherche, France, Programme ARPEGE LIG, LAMIH, LIRIS Service composition based framework for optimizing queries. Coordinator

CLEVER (2011 - 2013) STICAMSUD program U. de la República, Uruguay, UFRN Brazil, LIG, LAFMIA, LIFO, France Environment virtual observatory on cloud. Coordinator

SWANS (2014-2016) CNRS STiC-AMSUD Program U. de la Repblica, Uruguay, UFRN Brazil, LIG, LAFMIA, LIFO, France

(2013-2014) COST ICT Program : Semantic keyword-based search on structured data sources. Participant

SOGrid (2013-2017) ADEME Le réseau électrique de demain. Participant

SITAM (2012) Mexican Council of foreign affairs, Mexique Laboratorio Nacional de Informatica Aplicada, LAFMIA, UDLAP Summit Information Technologies in Mexico. Participant

AIWS (2012 - 2014) PEPS CNRS program (LIRIS-LIG) Discovering conversations among services by analyzing event logs - Coordinator

CAISES (2012 - 2015) European Union FP7, IRESES program (UK, France, Ukrania, China) Observation and industrial management on the cloud. Participant

S2EUNET (2010 - 2014) European Union FP7, IRESES program Norway, Spain, USA, Mexico, China Access to continuous data on heterogeneous networks. Participant

Key references:

- [1] V. Cuevas-Vicenttin, G. Vargas-Solar, C. Collet, Evaluating Hybrid Queries through Service Coordination in HYPATIA, In Proceedings of the 15th International Conference on Extending Database Technology (EDBT), Berlin, Germany, 2012
- [2] T. Delot, S. Ilarri, M. Thilliez, G. Vargas-Solar, S. Lecomte, Multi-scale query processing in vehicular networks, In Journal of Ambient Intelligence and Humanized Computing, Springer Verlag, ISSN 1868?5137, 2(3), 2011, pp. 213?226
- [3] Genoveva Vargas-Solar, Catarina Ferreira da Silva, Parisa Ghodous, José-Luis Zechinelli-Martini, Moving energy consumption control into the cloud by coordinating services, International Journal of Computing Applications, Special Issue. December 2013.
- [4] Carlos-Manual Lopez-Enriquez, Genoveva Vargas-Solar, José-Luis Zechinelli-Martini, Christine Collet, Hybrid query generation, LANMR, 11728,2012.
- [5] Valeria de Castro, Martin A. Musicante, Umberto Souza da Costa, Plácido A. de S ouza Neto, and Genoveva Vargas-Solar, Supporting Non-Functional Requirements in Services Software Development Process: An MDD Approach, In Proceedings of the 40th International Conference on Current Trends in Theory and Practice of Computer Science, LNCS Springer Verlag, High Tatras, Slovakia, January, 2014.
- [6] Javier A. Espinosa-Oviedo, Genoveva Vargas-Solar, José-Luis Zechinelli-Martini, Christine Collet. Policy driven services coordination for building social networks based applications. In Proc. of the 8th Int. Conference on Services Computing (SCC'11), Work-in-Progress Track, Washington, DC, USA, July 2011.

1.2.3 Reasoning on data semantics

This research is focused on combining reasoning and data management for efficiently querying and linking Web data through ontologies. Ontologies are very useful in many applications to express domain-specific knowledge over data that may be incomplete, uncertain or even inconsistent because coming from autonomous data sources distributed over the Web.

The proposed approach relies on recent complexity results showing that the expressive power of ontologies must be limited for making tractable reasoning on data enriched with ontologies. In particular, (several fragments of) the DL-Lite description logic we have been studied in the decentralized setting of P2P semantic networks. [1] designs a novel setting for robust module-based data

management allowing to re-use a part of a reference ontology-based data system as an independent module while guaranteeing that it evolves safely w.r.t both the reference schema and its associated data. We are investigating how it applies to extract modules from the knowledge base on anatomy of My Corporis Fabrica. [2,3] proposed a novel model of trust based on alignments between taxonomies for guiding the query answering process in P2P semantic networks. Finally, [4] provides a novel method that us systematic and mathematically well-founded for discovering mappings between taxonomies of classes.

Contracts: ANR projects: Continuum, Datarling, Qualinca, Pagoda

Key references:

- [1] Robust Module-based Data Management. Francois Goasdou and Marie-Christine Rousset. IEEE Transactions on Knowledge and Data Engineering , Volume 25, Issue 3, March 2013, pages 648-661
- [2] Alignment-based trust for resource finding in semantic P2P networks. Manuel Atencia, Jerome Euzenat, Giuseppe Pirro and Marie-Christine Rousset. Proceedings of ISWC 2011 (10th International Semantic Web Conference) .
- [3] Trust in Networks of Ontologies and Alignments. Manuel Atencia, Mustafa Al Bakri and Marie-Christine Rousset. Knowledge and Information Systems, to appear.
- [4] Discovery of Probabilistic Mappings between Taxonomies: Principles and Experiments Remi Tournaire, Jean-Marc Petit, Marie-Christine Rousset, and Alexandre Termier. Journal of Data Semantics (JoDS), Volume 15, pages 66-101.
- [5] Web Data Management, Serge Abiteboul, Ioana Manolescu, Philippe Rigaux, Marie-Christine Rousset, Pierre Senellart, book published by Cambridge University Press. description

1.2.4 Publications

Inserer Ici tableau rcapitulatif des publications

discussion de l'évolution des publications en terme de quantit et qualit entre le prcdent quadriennal et 2009-2014.

1.3 Visibility and attractivity

1.3.1 Rayonnement

** Membres IUF : - Marie-Christine Rousset ** Chevalier de l'ordre national du mrite *Marie – ChristineRousset(?), ChristineCollet(2010)*

Best Paper Awards

- Three prizes at SSSW 12 (Summer School on Ontology Engineering and the Semantic Web)
- Best Paper Award” de la track ”E”, embedded Software DATE 2014.

1.3.2 Principal International collaborations

China: Genoveva + Christophe corriger/completer SVP

Japon: Osaka University: A. Termier collaborates with Pr. Takashi Washio of I.S.I.R., Osaka University on graph mining algorithms.

NII: A. Termier also collaborates informally way with Takeaki Uno from NII, Tokyo on the parallelisation of Pr. Uno’s “LCM” algorithm. This collaboration has been formalized with the PhD student B. Nègrevergne (co-supervised by M.-C. Rousset and A. Termier).

Alex, corriger/completer STP

Mexique: the group has a long tradition in developing cooperations between the Mexican and French governments in TICS for 20 years. Since 2008, G. Vargas-Solar is deputy director of the French Mexican Laboratory in Informatics and Automatic Control (LAFMIA, UMI 371) an international unit of the CNRS. The cooperation of HADAS with Mexico has lead to scientific results and to the education of graduate students through co-advicing contracts (Orchestra project, ???) and the organisation of thematic schools. These activities include the participation of Ch. Collet and Ch. Bobineau and their PHD students The cooperation with Mexico includes the most important private and public institutions of that country: three major public research centres CINVESTAV, CICESE, INAOE, private centres like LANIA; and important universities like UDLAP, UATx, ITESM. The main research topics of the cooperation are services based infrastructures for managing distributed data with reliably, services based query processing and flexible data storage services. Collaboration on these topics have been formalized with PhD students: A. Portila ? , V. Cuevas, J. Espinosa-Oviedo, Carlos-Manuel LZ-ENREZ and Juan Carlos Castrejón.

– To be revised (Geno)

Vietnam: we have developed relations with the MICA laboratory in Hanoi (3 doctors formed) and participated to the Asian projects.

– To be completed (Christophe)

Brazil: Nous avons des collaborations avec l'Université Fédérale Rio Grande do Norte, département DIMAP, quipe FORALL depuis 2007. Les collaborations portent sur la gestion de données base de services sur le cloud: intégration sémantique de services de données travers les mashups, coordination de services base de politiques, traitement de données sur cloud avec des approches Map-Reduce, des langages d'interrogation de données base de composition de services. Nous avons privilégié des applications sur les données météorologiques et la distribution énergétique. Cette recherche s'est réalisée travers des projets soutenus par Microsoft (e-CLOUDSS), par le programme STICAMSUD (CLEVER et SWANS) et travers le co-encadrement d'étudiants de master et de thèse. A CAPES a également financé des postdocs et le CNRS des postes de prof invité.

???: other countries ???

– To be completed (all)

Contribution to the Scientific community

presidence of the conference BDA 2014: Ch. Collet

presidence of the EDBT 2013 school: Ch. Collet

Membre du conseil scientifique de la chaire d'excellence Smart Grids entre Grenoble INP et ERDF (2012-): Ch. Collet

Membre du comité de pilotage ANR, Modèles numériques (2010-2013): Ch. Collet

VP adjointe recherche groupe Grenoble INP: Ch. Collet (Avril 2007-2012)

Membre nommé du conseil scientifique INS2i - CNRS: Ch. Collet (2010-)

Chargée de Mission auprès du LIG pour la prospective scientifique: M.-C. Rousset

Membre du jury du prix de thèse Gilles Kahn (2010-2012) (prix décerné par Specif et patronné par l'Académie des Sciences): Ch. Collet

Elected president of the Mexican Society of Computer Science 2007-2009: G. Vargas-Solar

Deputy director of the UMI French Mexican Laboratory of Computer Science (LAFMIA, UMI 3571) (2008-) : G. Vargas-Solar

Management of Scientific Organisations

- *President of the Extended Database Technology (EDBT) association:* Ch. Collet (2013 -)
- *Member of the Extended Database Technology (EDBT) association:* Ch. Collet (2004 - 2013) in charge of the school programming
- *Member of the IJCAI-09 advisory committee:* M.-C. Rousset
- *Deputy director of the French Mexican Laboratory in Informatics and Automatic Control (LAFMIA, UMI 371) (2008 -) :* G. Vargas-Solar

Editorial Boards

- *PVLDB, publication of the Very Large Database Endowment*: Ch. Collet (2008-2010??)
- *Computacion y sistemas*, G. Vargas-Solar, since 2002.
- *e-Gnosis, electronic journal*, G. Vargas-Solar, since 2004.
- *ICDIM Journal special issue*, G. Vargas-Solar, since 2005.
- *KER Journal special issue*, G. Vargas-Solar, since 2007.
- *ActaPress Journal*, G. Vargas-Solar, since 2008.
- *Interstices*: M.-C. Rousset.
- *ACM Transactions on Internet Technology (TOIT)*: M.-C. Rousset, until 2005.
- *AI Communications*: M.-C. Rousset.
- *Communications of the ACM* : M.-C. Rousset, since 2009.

Organisation of Conferences and Workshops

- Extended Data Base Technology (EDBT) school 2009, Ch. Collet, T. Delot and G. Vargas-Solar
- Extended Data Base Technology (EDBT) school 2013, Sihem Amer-Yahia and G. Vargas-Solar

Program committee members

- *International Conference on Distributed Computing Systems*, Ch. Collet, 2007, 2009
- *Extended Data Base Technologies (EDBT)*, Ch. Collet, 2009,
- *International Conference on Data Mining (ICDM)*, A. Termier, 2009
- *SIAM International Conference on Data Mining (SDM)*, A. Termier, 2009
- *International workshop on ambient data integration (ADI)*, F. Jouanot, 2009
- *Bases de Données Avancées (BDA)*, Ch. Collet (2010, 2012, 2013), Ch. Bobineau (2009), G. Vargas-Solar(2011)
- *Gestion des Données dans les Systèmes d'Information Pervasifs (GEDSIP)*, Ch. Bobineau (2009), G. Vargas-Solar (2009)
- *International Workshop on Data and Services Management in Mobile Environments (D2SME)*, Ch. Bobineau (2009)
- *International ACM Conference on Management of Emergent Digital EcoSystems (MEDES)*, G. Vargas-Solar (2009)
- *International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, G. Vargas-Solar (2009)
- *ICPADS (International Conference on Parallel and Distributed Systems) track on "Web Services"*, G. Vargas-Solar, 2009

Je n'ai pas ré-actualiser - A COMPLETER par chacun

Evaluation committee members

- *Member of the evaluation panel for the ERC starting grants, Panel 5 (Information and Communication)*, M.-C. Rousset (2007-)
- *Member of the experts committee for projects evaluation of the LACCIR Microsoft virtual lab on TICS for Latin America*: G. Vargas-Solar (2007-)

National expertise

- *Vice-presidence comité d'évaluation scientifique ig Data, dcision, simulation, HPC h. Collet (2014-)*
- *ANR, Comit de pilotage Modles numériques, Ch. Collet, 2010-2013*
- *Member of the Specialists Council of the Delegation of Science and Technologies in Puebla, Mexico*., G. Vargas-Solar(2007-)

- A COMPLETER par chacun : ajouter les comités d'évaluation de labo, comités de recrutement

1.4 Social, economical, and cultural impact

Results of our research have direct impact on applications dealing with huge amounts of data and resources in pervasive environments. They include traditional enterprise applications such as mining logs, web applications but also "e-science" applications (in astronomy, biology, earth science, etc.). Environments we consider are wireless sensor networks (e.g. natural environment surveillance, industrial process monitoring), peer-to-peer data sharing, application deployment and maintenance for smartgrids, transports, networks ???

Application domains for data mining in structured data include: chemistry (e.g. molecule graphs), bioinformatics (e.g. gene or protein interaction networks), offline or online event logs mining.

For Social data ???

A REVOIR

1.4.1 Main Contracts and grants

Webcontent (RNTL, The semantic web framework-2006-2009), 10 partners(CEA LIST, EADS DCS, Thales Research & Technology, France Telecom R & D, ADRIA Développement, Soredab SAS, Exalead, New Phenix, Xyleme, INRIA-GEMO, INRA, INRIA-Mostrare, LIP6, PRISM, INRIA-InSitu, LIG, LIMSI-CNRS, GRIMM, EXMO, PSY-CO), (<http://www.webcontent-project.org/>), 83300 . Coordinator: CEA, Scientific lead in LIG: Ch. Collet and M.-C. Rousset. The WebContent project is creating a software platform to accommodate the tools necessary to efficiently exploit and extend the future of the Internet: the Semantic Web. The first targeted domain is the watch, a subpart of intelligence dedicated to warn the decider on the occurrence of an event or the evolution of a situation. It joins several Open Source tools to create

the core of a Service Oriented Application and it defines the interface of several services that are available through several partners, either freely or through commercial licences. These services then exchange data in a formalized manner.

OPTIMACS : (ANR, program ARPEGE 2008-2011), 3 partners (LIG, LAMIH, LIRIS), (<http://optimacs.imag.fr>), 227 128 . Coordinator: Grenoble INP-LIG, Scientific lead in LIG: G. Vargas-Solar. OPTIMACS (SERVICE COMPOSITION BASED FRAMEWORK FOR OPTIMIZING QUERIES) combines hybrid query processing and services composition, addressing services composition and query processing including adaptive hybrid query optimization according to QoS criteria. OPTIMACS is an original research project that will lead to results with an important expected impact on “modern data and services intensive systems” deployed on networks of heterogeneous devices, the so called ecosystem or dataspace.

DATARING (ANR, Programme Réseaux du futur et services 2008-2011), 3 partners (<http://www.lina.univ-nantes.fr/projets/DataRing/>). 130 549 . Coordinator: INRIA Nantes, Scientific lead in LIG: M.-C. Rousset. The DataRing project addresses the problem of P2P data sharing for online communities, by offering a high-level network ring across distributed data source owners. Users may be in high numbers and interested in different kinds of collaboration and sharing their knowledge, ideas, experiences, etc. Data sources can be in high numbers, fairly autonomous, i.e. locally owned and controlled, and highly heterogeneous with different semantics and structures. What we need then is new, decentralized data management techniques that scale up while addressing the autonomy, dynamic behavior and heterogeneity of both users and data sources.

CONTINUUM (ANR, Programme Réseaux du futur et services 2008-2011), 7 partners (I3S, LIG, SUEZ ENVIRONNEMENT, LYONNAISE DES EAUX, GEMALTO, LUDOTIC, MOBILEGOV), (<http://continuum.unice.fr>), 279 652 . Coordinator: University of Nice, Scientific lead in LIG: F. Jouanot and M.-C. Rousset. CONTINUUM (CONTINUITE DE SERVICE EN INFORMATIQUE UBIQUITAIRE ET MOBILE) addresses the problem of service continuity within the long-term vision of ambient intelligence. A core problem is to achieve software adaptation to a variety of resources in dynamic and heterogeneous environments with an appropriate balance between system autonomy and human control. Three key scientific issues will be addressed: context management and awareness, semantic heterogeneity, and human control versus system autonomy. The professions related to water management is used as a business application domain.

UBIQUEST (ANR, Programme BLANC 2009-2012), 3 partners (LIG, CITI, LIAMA), 149 099 . Coordinator: Grenoble INP-LIG, Scientific lead in LIG: Ch. Bobineau and Ch. Collet. UBIQUEST (Ubiquitous Quest: declarative approach for integrated network and data management in wireless multi-hop networks) aims at integrating network and data management in dynamic ad-hoc networks. this integration will be done by giving a distributed database view of the whole network. Each node stores network and application data in a local database. Messages between nodes

are queries or answers. The objective of this integration is the rapid development and deployment of applications and network protocols.

Sogrid (2013-2017) - ADEME Le réseau électrique de demain SoGrid aims to Confirm the path already opened by ERDF in the technological revolution of the Smart Grid. The goal is to develop the network intelligence by defining and creating all the components appearing in the chain of communication.

CASES 2012-2015) - European Union FP7, PEOPLE program (UK, France, Ukraine, China) Customised Advisory Services for Energy-efficient Manufacturing Systems. Coordinator

The project is aimed at teaming up trans-continental researchers in the areas of sustainable manufacturing and information technologies to enrich the knowledge base and achieve research synergies to develop smart design and manufacturing services in terms of energy efficiency. The Objectives are :

- 1) To establish an active international community and effective communication channel for research teams from various regions of the EU, China and Ukraine to collaborate in the research area of eco-design and energy efficient manufacturing planning.
- 2) To integrate the complementary expertise of the European, Chinese and Ukrainian teams to devise ICT-based smart services and standards to address the multi-faceted requirements of global eco-design and sustainable manufacturing planning.
- 3) To leverage the geographical characteristics of all teams for research evaluation, improvement and dissemination.

Smart Energy (2012-2014) Groupe de travail sur les Smartgrids; Participants : Grenoble INP LIG, G-SCOP, G2ELab, Gipsa-Lab

CLEVER (2012-2013) CLEVER CLOUD-BASED LATIN-AMERICAN ENVIRONMENTAL VIRTUAL OBSERVATORY).

The project aims at providing the underlying services that will enable the VO to personalize and manage mashed up services. The result will be a platform where climate reports coming from different providers in LATAM will be mashed up. Resulting mashups will be exported as VO tools for eventually building other mashups.

QUALINCA (2012-2015)

QUALINCA is a ANR Contint funded research project looking at developing mechanisms allowing to quantify the quality level of a bibliographical knowledge base, to improve the afore mentioned quality level, to maintain the quality when updating the knowledge base and to exploit the knowledge bases taking into account their quality levels. This project aims to develop mechanisms to:

- describe the quality of an existing document database;
- maintain a given level of quality by controlling updates on such databases;
- improve the quality of a database;

- exploit these databases according to their level of quality.

SocTrace (2011-2015) FUI-Minalogic, OSEO 2011-2015

Partners: INRIA, LIG, TIMA, STMicroelectronics, Magilem, probayes

Coordinator: STMicroelectronics

The SoC-Trace project aims to develop a set of methods and tools based on traces of execution produced by multi-core embedded applications. It will allow developers to optimize and debug these applications more efficiently. Such methods and tools should become a building block for the design of embedded software, in response to the growing needs of analysis and debugging required by the industry. The technological barriers consist of a scaling problem (millions of events stored on gigabytes) and a trace understanding problem related to applications whose complexity is increasing. The project addresses the problem of controlling the volume of tracks and of developing new analysis techniques. SocTrace is composed of academic partners with related themes, and several industry partners including STMicroelectronics.

1.4.2 Research Networks (European, National, Regional, Local)

E-CLOUDS: (BUILDING E-GOVERNEMENT CLOUDS USING DISTRIBUTED SEMANTIC SERVICES, Microsoft, 2007-2011, LACCIR, <http://e-cloudss.imag.fr>), 5 partners (CNRS LIG-LAFMIA, Fundacion Universidad de las Américas, Puebla, Mexique, Universidad de la Republica de Uruguay, Uruguay, Universidade Federal do Rio Grande do Norte), 50,000 USD. Coordinator: J.L. Zechinelli Martini, LAFMIA, Scientific lead in LIG: G. Vargas-Solar). The objective of E-CLOUDSS is to propose an infrastructure for mashing up reliable semantic services for building e-government clouds. Mashups represent a new wave for building Web applications. E-CLOUDSS addresses the management (definition and enforcing at execution time) of non functional properties associated to services' coordination for building reliable mashups. Effective ways to perform virtual executions is one of the main subjects of study of E-CLOUDSS.

WebIntelligence: (Cluster Régional "Informatique, Signal, Logiciels embarqués" - 2006-2009). The project aims at organizing research on web intelligence in Rhone-Alpes.

ORCHESTRA: (ORCHESTRATION TRANSACTIONNELLE DE SERVICES, Program: ECOS-ANUIES 2007-2011), 3 partners (Grenoble INP, Universidad Autonoma de Tlaxcala, Fundacion UNiversidad de las Américas, Puebla, Mexique). Missions for Professors (Ch. Collet in 2007 and 2008, and G. Vargas in 2009) and PhD students. The objective of ORCHESTRA is to propose an infrastructure pour building transactional, secure and evolutive service-based applications. The key elements of the project are: (i) the definition of a framework (general solution) of technical services for managing the security, transactional properties and evolution of business services ; and (ii) implementation of the framework an its validation in the development of service-based applications: production chains.

1.4.3 Internal Funding

RED-SHINE: (RELIABLY AND SEMANTICALLY INTEGRATING WEB INFORMATION BY MASHING UP DATA SERVICES, BQR Grenoble INP, 2009). 2 partners (LIG, LAFMIA-UMI 3175) (<http://lafmia.weebly.com/>), 20 000- one PhD grant and 4 months for inviting professors. Coordinator: Grenoble INP-LIG, Scientific lead in LIG: G. Vargas-Solar).

The objective of RED-SHINE is to propose an infrastructure for mashing up services using semantics and thereby integrating information from the Web. RED-SHINE will redefine and extend OQLiST for declaratively defining reliable semantic mashups. RED-SHINE addresses the management (definition and enforcing at execution time) of non functional properties (NF-P) associated to services' coordination for building reliable mashups. The objective of our work will be to propose a language for orthogonally expressing NF-P and ensuring strategies, and to specify execution strategies for adding NF-P to mashups.

DAMOCLES: (MSTIQ project, 2009). 2 partners (LIG, TIMA), 15 000 - one year postdoc. Coordinator: Grenoble INP-LIG, Scientific lead in LIG: A. Termier).

DAMOCLES (DAta Mining for On Chip Low Energy Systems) aims at developing data mining algorithms for analysing memory accesses in System-on-Chip processors, in order to optimise data placement and thus reduce energy consumption.

smart Energy: Grenoble INP Project (2012-2015) - LIG, G-SCOP, G2ELab, Gipsa-Lab

WalT: Grenoble INP niversit Joseph Fourier, Programme AGIR (2013-2015) on Wireless Testbed

1.5 Team Organization and life

1.6 Training through research, educational involvment

Sur la priode 2009-2014:

Thesis: 11 thesis have been defended; - On average, the thesis is done in four years.

Suivi: regular meetings, presentations of paper, "workshop" on talks / papers, participation to schools,

Devenir: Le devenir des doctorants est vari : X dans l'industrie (Chercheur ou ingnieur RD) ; 2 dans le suprieur. The becoming of the students is varied: X are in industry (researcher or engineer R D) , YY have academic positions

Educational involvment

1.7 Strategy and Research Project

1.8 Self assesment