

# Curriculum Vitæ

## Stéphane Gancarski

23 novembre 2010

Né le 12 mai 1965 à Strasbourg. Marié, 2 enfants. Nationalité française.  
Section CNU : 27. Titulaire de la PEDR depuis 2000.  
Qualifié aux fonctions de Professeur en 2006  
Maître de Conférences à l'Université Pierre et Marie Curie Paris 6  
Coordonnées : LIP6, 4 place Jussieu, 75252 PARIS cedex 5  
tél : 01 44 27 87 95 – fax : 01 44 27 70 00  
email : Stephane.Gancarski@lip6.fr  
homepage : www-poleia.lip6.fr/~gancarsk

Références :

Anne Doucet, Professeur UPMC, Anne.Doucet@lip6.fr

Patrick Valduriez, Directeur de Recherches INRIA, Patrick.Valduriez@inria.fr

## Table des matières

<b>1</b>	<b>Formation</b>	<b>2</b>
<b>2</b>	<b>Parcours Professionnel</b>	<b>2</b>
<b>3</b>	<b>Synthèse de la carrière</b>	<b>2</b>
<b>4</b>	<b>Enseignement</b>	<b>3</b>
4.1	Résumé . . . . .	3
4.2	Bilan personnel et motivations . . . . .	3
4.3	Présentation synthétique des enseignements . . . . .	4
<b>5</b>	<b>Activités scientifiques</b>	<b>4</b>
5.1	Récapitulatif des encadrements et publications . . . . .	4
5.2	Responsabilités, instances scientifiques, jurys (hors projets et coopérations) . . . . .	5
5.2.1	Commissions . . . . .	5
5.2.2	Comités de programmes, organisation scientifique et relectures . . . . .	5
5.2.3	Rapports et jurys de thèse . . . . .	5
5.2.4	Organisation, coordination et expertise . . . . .	6
5.3	Projets et coopérations . . . . .	6
<b>6</b>	<b>Description des recherches depuis 2000</b>	<b>6</b>
6.1	Contraintes d'intégrité et transactions imbriquées (depuis 2000) . . . . .	7
6.2	Fouille de données dans les entrepôts de données (2000) . . . . .	7
6.3	Optimisation de performances par relâchement de la cohérence et réplication sur un cluster de PC (2001-2007) . . . . .	7
6.4	Transactions et mobilité dans les bases de données (2002-2004) . . . . .	8
6.5	Partage de ressources scientifiques par médiation (2002-2005) . . . . .	9
6.6	Gestion de données en Pair-à-pair (depuis 2006) . . . . .	9
6.7	Archivage du Web et préservation numérique (depuis 2008) . . . . .	10
6.8	Perspectives . . . . .	11
<b>7</b>	<b>Liste de publications</b>	<b>11</b>

## 1 Formation

- 2006** Habilitation à Diriger les Recherches en Informatique à l'Université de Paris 6  
Laboratoire : LIP6, Université de Paris 6. Mention : Très Honorable.  
Titre : *Cohérence et fraîcheur dans les bases de données réparties.*  
Jury : P. Pucheral (rap.), M. Scholl (rap.), D. Shasha (rap.), A. Doucet, M. Rukoz, P. Valduriez.
- 1994** Thèse de Doctorat en Informatique à l'Université de Paris 11 Orsay  
Laboratoire : Lamsade, Université Paris-Dauphine. Mention : Très Honorable.  
Titre : *Versions et Bases de Données : modèle formel, supports de langage et d'interface-utilisateur*  
Jury : M. Adiba (rap.), F. Bancilhon, N. Bidoit, W. Cellary (rap.), C. Delobel, A. Doucet, G. Jomier (dir.)
- 1991** DEA d'Informatique. Université de Paris 11. Mention Bien.
- Antérieur** DESS Université Paris-Dauphine (1990). ENSAE, Ecole Nationale de la Statistique et de l'Administration Economique (1985-87).
- Langues** : Anglais, portugais. Notions d'espagnol et d'allemand.

## 2 Parcours Professionnel

- depuis 1996** Maître de Conférences à l'Université Pierre et Marie Curie Paris 6.
- 1995-96** A.T.E.R. à l'Université Paris 12 Créteil.
- 1994-95** A.T.E.R. à l'Université Paris 11 Orsay.
- 1991-94** Allocataire-Moniteur à l'Université Paris-Dauphine.
- Antérieur** Analyste - programmeur, Concepteur PAO.

## 3 Synthèse de la carrière

A mon arrivée à l'UPMC en 1996 (après ma thèse et deux années d'ATER), les besoins en enseignement en bases de données étaient énormes. L'équipe était alors composée de deux personnes, et beaucoup d'enseignements à redéfinir et assurer, pour des promotions de plus de 300 étudiants. C'est pourquoi je me suis particulièrement investi dans l'enseignement, en assurant les cours magistraux (licence et maîtrise) et la gestion d'équipe pédagogique, en rédigeant un polycopié de 60 pages (pour les étudiants, mais aussi les collègues non spécialistes).

Lorsque l'équipe s'est agrandie, j'ai pu consacrer plus de temps à la recherche, en mettant en œuvre une collaboration (financée et reconduite plusieurs fois par le CNRS) avec l'équipe « parallélisme et répartition » de l'université centrale du Venezuela à Caracas. Cette collaboration a non seulement donné lieu à plusieurs séjours de part et d'autres, de nombreuses publications, mais aussi à une thèse en cotutelle que j'ai encadrée côté français. Cette collaboration continue, puisque je devrais accueillir en cotutelle un nouveau doctorant de cette équipe, suite à un stage dans notre équipe l'an dernier.

Avec l'émergence des financements sur projet, j'ai pu prendre de nouvelles responsabilités. Tout d'abord localement, dans le cadre du projet RNTL Leg@net où j'ai animé une équipe de 3 personnes (dont une thèse que j'encadrais), puis globalement puisque j'ai été porteur du projet ANR Respire qui réunissait 5 équipes de recherche. Respire a fédéré une bonne partie de la recherche en France sur la gestion de données en pair à pair, donnant lieu à des collaborations internes mais aussi avec d'autres projets connexes. C'est d'ailleurs dans le cadre de Respire que j'ai organisé plusieurs Workshops, dont l'un associé à la conférence EDBT, l'une des conférences majeures du domaine.

Ces différentes responsabilités m'ont progressivement intégré à la communauté internationale. J'ai ainsi été convié à faire partie de différents comités de programmes, dont les conférences internationales prestigieuses IEEE ICDCS et ICDE, mais aussi à faire partie du comité scientifique de la conférence portugaise XATA. Mon intégration dans la communauté nationale date, elle, de mes années de doctorat, notamment par mon implication dans le GDR. Je suis régulièrement convié à faire partie du comité de programme de la conférence BDA qui fédère la communauté depuis près de 30 ans. Par ailleurs, mon activité dans *Respire* m'a associé aussi à la communauté «systèmes répartis», la confluence des deux thématiques étant d'ailleurs une caractéristique de *Respire* et d'une grande partie de mes travaux.

## 4 Enseignement

### 4.1 Résumé

Ma carrière d'enseignant a commencé par un poste d'allocataire-moniteur à l'Université Paris-Dauphine et deux années d'ATER, à l'Université Paris 11 Orsay puis à l'Université Paris 12 Créteil. A ma nomination à l'Université Pierre et Marie Curie Paris 6, mon enseignement s'est tout d'abord concentré dans l'enseignement des bases de données, puisque non seulement j'ai été le responsable de l'enseignement de bases de données en licence de 1996 à 2004, mais j'ai eu aussi certaines années la responsabilité des cours de bases de données en maîtrise et en MST (première et deuxième année). De plus, je me suis investi en DEA, en groupe de recherche tout d'abord, en cours de tronc commun depuis 2002. De ce fait, j'ai acquis une grande expérience dans l'enseignement des bases de données, à tous les niveaux de formations et sur presque tous les sujets liés aux bases de données, des aspects modèle/langage (SQL, OQL, SQL3, XML) jusqu'aux aspects les plus proches du système. Ceci m'a permis de participer activement à la définition de la plupart des modules de bases de données enseignés à l'UFR d'Informatique de mon Université. En tant que responsable de l'enseignement des bases de données en licence, j'ai animé et coordonné pendant huit ans une équipe de chargés de TD (jusqu'à 12 groupes). Un point central de cette coordination a été de rédiger un polycopié de 60 pages [Gan03] pouvant servir de référence aussi bien aux étudiants qu'aux enseignants débutant dans la matière. Il est à noter que ce polycopié est toujours largement utilisé.

Avec l'arrivée de nouveaux collègues spécialistes de bases de données, j'ai eu l'opportunité de me diversifier dans d'autres disciplines, principalement en réseaux où j'ai assuré des TD-TP en licence de 2000 à 2004.

Enfin, j'ai aussi eu l'occasion de donner des cours à l'extérieur de mon Université, notamment en DEA à l'Université Paris-Dauphine, à l'Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE), à l'Institut National des Télécommunications (INT), ainsi qu'en Master à l'ESSEC.

### 4.2 Bilan personnel et motivations

Comme il est écrit précédemment, j'ai assumé beaucoup de responsabilités d'enseignement dès le début de ma carrière de Maître de conférence (j'ai même assumé la responsabilité de tous les cours de bases de données à l'UFR d'Informatique de l'UPMC pendant un an). Assurant aussi certains groupes de TD de ces cours, j'ai ainsi pu remettre à jour de manière cohérente les programmes de ces différentes modules en introduisant de nouveaux sujets tels les vues, les contraintes, l'évaluation de requêtes, la répartition et la réplication des données etc. La cohérence des enseignements concerne aussi bien le lien entre cours et TD/TP, mais aussi le lien entre les différentes années de la formation. J'ai aussi introduit à l'UFR l'apprentissage de la pratique des SGBD par des séances de travaux sur machine.

Enfin, plus récemment, j'ai participé à de nouveaux enseignements visant à faire le lien entre les nouvelles technologies du Web (XML et ses langages, PHP, etc) et les bases de données.

Cette grande expérience me permet d'être opérationnel assez rapidement pour tout enseignement (existant ou à créer) dans le domaine des bases de données. De plus, mes compétences de chercheur me motivent à proposer des enseignements plus spécifiques au niveau M, à la frontière entre les bases de données et les systèmes répartis (et si possible, en coordination avec les enseignants spécialistes de systèmes répartis). Par exemple, un module pourrait présenter les différents aspects de la cohérence en environnement réparti : répartition et réplication des données, protocoles de cohérence entre répliques, modèles transactionnels, cohérence sémantique. Pour chacun de ces aspects, on présente dans un premier temps les aspects conceptuels qui permettent de définir la cohérence, puis dans un second temps les impacts en termes de performances des différentes solutions possibles. En effet, la cohérence et les performances sont pour moi deux composantes primordiales de la qualité de service d'un système d'information. Il me paraît donc très important (et par conséquent motivant en tant qu'enseignant) de faire comprendre aux étudiants que ces deux composantes sont le plus souvent conflictuelles et qu'il est nécessaire de bien en appréhender les interactions pour trouver le compromis le mieux adapté.

De manière plus générale, j'ai toujours eu à coeur de proposer des enseignements équilibrés entre théorie et pratique. Comme pour une langue étrangère, une théorie est mieux assimilée si son utilité est démontrée par la pratique. Par exemple, faire programmer des triggers, parfois lourds, pour maintenir des dépendances fonctionnelles permet de motiver les étudiants à la théorie de la normalisation des schémas relationnels. Par ailleurs, leur faire écrire et exécuter des requêtes aux multiples jointures sur des schémas très éclatés permet de leur faire appréhender les limites de la même théorie.

Enfin, mes motivations pour l'enseignement et son organisation ne se limitent pas à ma spécialité. Si nécessaire, je suis prêt à prendre en charge tout enseignement ou responsabilité pour lesquels j'aurais les compétences nécessaires.

### 4.3 Présentation synthétique des enseignements

Depuis quelques années, mon service d'enseignement est assez stable, et couvre une bonne partie des enseignements en bases de données de l'UFR. Bien qu'ayant déjà enseigné en formation professionnelle, j'enseigne à l'heure actuelle uniquement en formation initiale devant les étudiants. Globalement, la répartition est la suivante :

- TD en L3 : 40h
- TD en M1 : 80h
- Cours de L3 : quelques heures par an
- Cours en M1 : 20h
- Cours en M2 : 20h
- Encadrement de projets : 10h

De plus, j'encadre chaque année un ou deux stagiaires de M2.

## 5 Activités scientifiques

### 5.1 Récapitulatif des encadrements et publications

- 4 thèses de doctorat en encadrant principal
  - ◇ Claudia Leon. *Contraintes d'intégrité et transactions imbriquées*. 2001.  
Fonction actuelle : Maître de conférences à l'Université de Caracas.
  - ◇ Cécile Le Pape. *Contrôle de Qualité des Données Répliquées sur un Cluster*. 2005.  
Fonction actuelle : Maître de conférences à l'Université de Paris 6.
  - ◇ Myriam Ben-Saad. *Archivage du Web : complétude et cohérence*. En cours depuis septembre 2008.

- ◇ Zeynep Pehlivan. *Recherche d'information dans les archives Web*. En cours depuis octobre 2009.

Par ailleurs, j'ai participé d'assez près à l'encadrement de la thèse d'Idrissa Sarr, soutenue en octobre 2010 et dont l'encadrant principal est Hubert Naacke.

- 16 DEA/M2 et 1 DESS.
- 45 articles avec comité de sélection dont 4 en revues internationales, 5 en revues nationales, 2 en revues internationales en ligne, 25 en conférences internationales (dont un article court) et 9 en conférences nationales (voir liste de publications).
- 2 chapitres de livre (voir liste de publications).

## **5.2 Responsabilités, instances scientifiques, jurys (hors projets et coopérations)**

### **5.2.1 Commissions**

- Membre de 2000 à 2005 de la commission de spécialistes 27<sup>e</sup> section, Université Paris 6. Membre extérieur nommé à l'Université Paris 5 en 2003. Membre extérieur suppléant au CNAM Paris depuis 2007.
- Membre nommé (titulaire, mais actuellement suppléant) du conseil scientifique du LIP6.
- Membre élu du conseil de laboratoire du LIP6 jusqu'en 2007.
- Membre élu du conseil de l'UFR d'Informatique de l'Université de Paris 6 de 2005 à 2007. Membre suppléant de la commission du personnel.
- Expert pour l'évaluation d'un projet RNTL dans le cadre de l'appel 2006. Expert pour les projets ANR depuis 2008.

### **5.2.2 Comités de programmes, organisation scientifique et relectures**

- Program chair et organisateur du Workshop international DaMaP (Data Management in Peer-to-peer systems), joint à la conférence internationale EDBT'08, Nantes. Organisateur de deux workshops sur invitation dans le domaine de la gestion de données en pair-à-pair (Paris, février 2007 et Le Croisic, juin 2007).
- Exhibit Chair de la conférence internationale SIGMOD/PODS en 2004.
- Membre du comité de programme des conférences internationales CoopIs (2004), IEEE ICDE (2007 et 2008) et IEEE ICDCS (area Data Management, 2007).
- Membre du comité de programme des conférences nationales
  - ◇ BDA (Bases de données avancées) en 1997, 1999, 2000, 2003, 2005, 2008.
  - ◇ UbiMob (Ubiquité et Mobilité) en 2006, 2008 et 2009.
  - ◇ CDUR (Cohérence des Données en Univers Réparti) 2006 et 2009.
- Membre du comité scientifique de la conférence portugaise XATA (XML, applications et technologies associées) depuis 2007.
- Rapporteur pour la revue internationale TAPOS (Theory and Practice of Object Systems) en 2000. Relecteur extérieur pour les conférences internationales VLDB 2005 et PODS 2005. Relecteur en 2010 pour la revue internationale Data and Knowledge Engineering.

### **5.2.3 Rapports et jurys de thèse**

- Rapporteur de 3 thèses : Antonio Souza, Université du Minho, Portugal, 2007, Samy Semghouni, Université du Havre, 2007 et Gilles Dodinet, Université Paris-Dauphine, 2010.
- Membre du jury de 10 thèses (hors celles que j'ai encadrées et rapportées) : Maude Manouvrier, Université Paris-Dauphine, 2000, Nicolas Lumineau, Université de Paris 6, 2005, Cédric Coulon, Université de Nantes, 2006, Corina Ferdean, Université de Paris 6, 2006, Reza Akbarinia, Université de Nantes, 2007, Khaled Jouini, Université de Paris-Dauphine, 2008, Rabab

Hayek, Université de Nantes, 2009, Loïc Cudennec, Université de Rennes, 2009, Anneli Lenica, Université P. et M. Curie 2010, Pierre Sutra, Université Paris-Dauphine, 2010.

#### 5.2.4 Organisation, coordination et expertise

- Secrétaire scientifique de la revue I3 (Cepaduès éd.) (2001/2002). Organisation d'un séminaire sur la Recherche d'Information en janvier 2002.
- Expert pour l'évaluation d'un projet RNTL dans le cadre de l'appel 2006.
- Organisateur du Séminaire Bases de données de l'Île de France, 2000.
- Coordonateur de l'axe «Cohérence» de l'action spécifique (AS) «Mobilité / Accès aux données» du CNRS (2002/2003)

#### 5.3 Projets et coopérations

- Responsable pour le LIP6 du projet européen SCAPE (FP7 - IP), dont le démarrage est prévu en février 2011. Responsable d'un workpackage sur la qualité des archives.
- Responsable pour le LIP6 du projet CARTEC financé par l'ANR sur l'archivage du Web depuis janvier 2008. Les deux autres partenaires sont l'INA (Institut National de l'Audiovisuel) et la société Antidot, spécialiste des moteurs de recherches et de l'indexation full-text.
- Responsable scientifique du projet RESPIRE, financé par l'ANR (projets ARA Masse de données), regroupant le LIP6, le LINA (Nantes), l'IRISA (Rennes), l'INT (Evry) et l'INRIA (Rocquencourt). Présentation du projet aux journées PariStic Nancy le 24 novembre 2006 (voir sur paristic.loria.fr). Revue finale en novembre 2008, très bonne évaluation.
- Responsable scientifique en France du projet de collaboration CNRS-FONACYT (équivalent vénézuélien du CNRS) entre le LIP6 et le CCPD de l'Université Centrale du Venezuela, Caracas (renouvelé chaque année de 2000 à 2004).
- Membre du projet CARTEC (ANR MDCO, depuis fin 2007), regroupant le LIP6, l'INA et la société Antidot, autour de l'archivage du Web et son suivi temporel.
- Membre du projet SEMWEB (ACI MD, 2004-2007), regroupant le LIP6, le CNAM (Paris), le LINA (Nantes), le LIRIS (Lyon), le LSI (Toulon) et le PRISM (Versailles).
- Membre du projet RNTL Leg@net, regroupant le LIP6 et les sociétés Prologue et ASPLine (2001-2003). Responsable pour le LIP6 au départ de P. Valduriez en 2002.
- Membre du projet PADOUE (ACI GRID, 2001-2005), regroupant le LIP6, le LIRMM, le CEMAGREF, l'INRIA, l'IRD-ORSTOM et le CDS (Strasbourg).

## 6 Description des recherches depuis 2000

Mes recherches concernent presque toutes une problématique importante dans les bases de données réparties : comment trouver le meilleur compromis entre la cohérence et les performances, autour du concept central de transaction ? La cohérence, qui reflète la qualité des données et des traitements se décline suivant plusieurs dimensions, telles que l'intégrité, la cohérence mutuelle des copies et des versions, l'isolation et l'atomicité. Les performances sont de deux natures : rapidité du système (temps de réponse) et économique (moindre coût de développement et d'adaptation). Après les travaux liés à ma thèse [GJ94, GJ95, GJ96, GJ01, CGJM01] sur la cohérence des version, et les travaux connexes qui ont immédiatement suivi [DGJM96, DGJM96a, DGJM96b, DFGJM96, GRW97] sur l'intégrité des bases à versions, j'ai continué à explorer ce compromis en considérant la réplication, les modèles de transactions avancés et plus récemment, l'archivage du Web. Ces travaux ont porté sur différents types d'environnements répartis : multibase, grappe de PC, grille, mobile et pair-à-pair, et ont eu pour cadre différents projets (RNTL, ACI, ANR-ARA, collaboration internationale). Par ailleurs, les collaborations internes au LIP6 m'ont amené à aborder d'autres sujets de recherche liés aux bases de données, comme la localisation grâce aux métadonnées ou l'exploitation de bases

de données multidimensionnelles pour la fouille de données. Tous ces travaux sont présentés dans cette section par ordre chronologique et suivis par mes perspectives de recherche.

## **6.1 Contraintes d'intégrité et transactions imbriquées (depuis 2000)**

Les contraintes d'intégrité sont le moyen le plus courant pour assurer la cohérence sémantique dans les bases de données. L'un des principaux problèmes de l'utilisation de contraintes d'intégrité est que, lorsque l'une d'entre-elles est violée par une transaction, cette dernière doit être abandonnée et donc défaire entièrement. Ce problème est particulièrement crucial dans certaines applications, conception notamment, où les transactions peuvent être de très longue durée et où il est très coûteux de les défaire. Les modèles de transactions imbriquées pallient partiellement cet inconvénient en proposant de découper une transaction en une hiérarchie de morceaux (sous-transactions) de taille raisonnable et indépendants, qui peuvent s'exécuter et être abandonnés séparément. Notre problématique est d'adapter la gestion de contraintes d'intégrité dans le cadre de telles transactions, en essayant de vérifier les contraintes dès lors que les sous-transactions susceptibles de les violer ont toutes fini de s'exécuter, mais sans attendre la fin des autres sous-transactions. Après avoir mis au point une première méthode résolvant ce problème, nos travaux ont ensuite visé à l'étendre dans le cas où transactions et contraintes sont définies sur une base de données répartie. Par ailleurs, nous avons appliqué en partie ces travaux dans le cadre d'une application de commerce électronique, définissant un agent intermédiaire pour effectuer des transactions sûres sur l'Internet. Il s'est avéré dans ce contexte qu'un modèle de transactions imbriqué «ouvert» serait mieux adapté car il permet de relâcher l'atomicité et l'isolation que le modèle «fermé» utilisé jusque là.

Ce travail a été mené jusqu'en 2005 dans le cadre d'une collaboration CNRS-FONACYT (équivalent vénézuélien du CNRS) dont j'étais le responsable côté français, entre le LIP6 et le CCPD de l'Université Centrale du Venezuela, Caracas. Il a abouti à plusieurs publications [DGLR01, GLR00, DGLR00, DGLR00, DGLLR02, GLNRS05, GLNRS06] et plusieurs prototypes. Il a donné lieu à la thèse en cotutelle de Claudia Leon que j'ai co-encadrée au LIP6 ainsi qu'au Maestria (MSc) de Rainell Grau que j'ai partiellement encadré lors de sa venue au LIP6 et lors de mes missions à Caracas.

## **6.2 Fouille de données dans les entrepôts de données (2000)**

L'extraction de connaissances à partir de données (Knowledge Discovery from Database (KDD)) est actuellement un domaine très actif dans la communauté informatique, et les entrepôts de données constituent une cible privilégiée pour les algorithmes d'apprentissage. Mais ceux-ci ne sont pas toujours adaptés à travailler sur de telles masses de données. Par l'intermédiaire, des bases de données multidimensionnelles une grande partie de la préparation des données peut cependant être effectuée. Notre travail a été de rechercher les apports de ces bases de données multidimensionnelles pour les algorithmes de fouille de données, c'est-à-dire d'étudier comment l'information agrégée, telle qu'elle est présentée dans de telles bases, permet de faire de l'extraction de connaissances correcte.

Ces travaux ont été menés en collaboration avec l'équipe «logique floue et traitement d'incertitude» du LIP6 dans le cadre d'un projet agréé et financé par le LIP6. Ils ont donné lieu notamment au stage de DEA d'Anne Laurent, co-encadré avec Christophe Marsala et à deux articles [LBDGM00, LGM00], ainsi qu'un chapitre de livre sur les bases de données multidimensionnelles [DG01]

## **6.3 Optimisation de performances par relâchement de la cohérence et réplication sur un cluster de PC (2001-2007)**

Les grappes de PC (PC cluster) sont le moyen le plus économiquement abordable pour bénéficier des avantages des machines parallèles : équilibrage de charge et robustesse aux pannes. Elles sont au cœur de la stratégie économique des hébergeurs d'applications (ASP en anglais). Dans le modèle

ASP, le client désire que son application existante soit transférée sur le cluster d'un hébergeur afin de se délivrer de tâches fastidieuses de maintenance (ex. sauvegarde) et accéder à ses données via l'internet. Cet hébergement, pour être viable, doit se faire au moindre coût tout en gardant des performances au moins aussi bonnes. Ce point est un des objectifs principaux du projet RNTL Leg@net, regroupant le LIP6 et les sociétés Prologue et ASPLine dans le cadre duquel a démarré cette recherche. La solution évidente pour obtenir de bonnes performances est de répliquer la base de données sur plusieurs nœuds du cluster de l'hébergeur. L'hébergement au moindre coût doit se faire avec un minimum d'adaptation lors du portage sur le cluster de l'hébergeur : les applications, les bases de données doivent rester inchangées. De même, il est indispensable d'utiliser un SGBD du marché et si possible de ne pas utiliser les fonctionnalités internes de celui-ci en terme de réplication car elle nécessiterait une adaptation trop importante du code existant des applications. C'est pourquoi nous avons choisi une architecture intergicielle : notre système intercepte les appels SQL des applications, calcule le nœud d'exécution optimisant l'équilibrage de charge et lui envoie l'appel SQL de l'application.

Ces dispositions architecturales sont détaillées dans les premiers articles que nous avons rédigés sur le sujet [GNV02, GNPV02]. Un deuxième point abordé dans ces articles est que d'autres gains de performances peuvent être obtenus en relâchant la cohérence selon les besoins de l'utilisateur et en exploitant les connaissances sémantiques sur l'application pour améliorer le degré de concurrence dans l'exécution parallèle sur plusieurs nœuds du cluster. Dans ce contexte, mes activités se sont concentrées sur deux points particuliers :

- Mode mono-maître : dans ce mode, un seul nœud (nœud maître) accepte les mises à jour et la cohérence est donc réglée par le gestionnaire de transactions local à ce nœud. Les autres nœuds, esclaves, sont rafraîchis de manière asynchrone et servent à l'exécution des requêtes en lecture seulement. Dans le cadre du stage de DEA puis de la thèse de Cécile Le Pape que j'ai encadrée, nous avons cherché à exploiter la sémantique de ces requêtes afin d'optimiser leur exécution sur les nœuds esclaves. Pour cela nous exploitons principalement deux dimensions : la fraîcheur et la granularité. Grâce à un modèle assez simple, les utilisateurs peuvent définir le niveau de fraîcheur qu'ils désirent pour le résultat de leur requête, à un niveau de granularité variable (la base de données, la relation ou l'attribut). Le système choisit ensuite le meilleur nœud pour exécuter la requête en calculant, pour chaque nœud, s'il est assez frais par rapport à la requête et sinon, quel est le coût de le rafraîchir. Le modèle de fraîcheur a donné lieu à deux articles [GLV03, LGV03]. Afin de ne pas laisser les nœuds devenir trop obsolètes, des procédures de rafraîchissement indépendantes du routage ont été ajoutées au système. L'implémentation a été démontrée lors de la revue du projet RNTL et les résultats expérimentaux très satisfaisants ont été publiés dans [LGV04, LG06].
- Mode multi-maître : ce mode est une extension au mode mono-maître, permettant que les transactions soient exécutées sur n'importe quel nœud. La grande différence est que les transactions sont partiellement ordonnées sous forme d'un graphe global. Le rafraîchissement d'un nœud consiste donc ici à exécuter des sous-graphes du graphe global. Architecture et modèles ont été publiés dans [GLN05]. Les résultats obtenus grâce au rafraîchissement dépendant du routage ont été publiés dans [GNPV06].

## 6.4 Transactions et mobilité dans les bases de données (2002-2004)

On définit un environnement mobile comme étant un environnement constitué d'une ou plusieurs unités mobiles (ex : PDA, portables) et éventuellement d'une ou plusieurs unités fixes. Les ressources des unités mobiles sont souvent assez limitées en mémoire, en capacité de stockage, en bande passante et en puissance de batterie. Ces diverses caractéristiques compliquent la gestion de base de données et restreignent la qualité des services proposés par les applications. Plusieurs raisons expliquent pourquoi les transactions classiques ne peuvent pas être exécutées dans un environnement mobile. Les déconnexions fréquentes des unités mobiles affectent la phase importante de validation de la



transaction. L'attente prolongée pour une re-connexion éventuelle de l'unité mobile fait échouer la validation de toute la transaction car les ressources ne peuvent être bloquées (verrouillage) sur le serveur fixe indéfiniment.

Ce problème explique le nombre important et croissant de travaux sur les transactions en milieu mobile. Dans un premier temps, nous nous sommes attaché à faire une synthèse de ces travaux, puis à proposer une nouvelle approche qui se base sur la notion de transaction souple de longue durée et sur la notion de contrat. Le contrat entre une unité mobile et le serveur fixe exprime les contraintes d'exécution (délai à respecter, restrictions sur les opérations possibles) que doit respecter l'unité mobile afin que le serveur fixe lui offre des garanties sur la validation finale du travail effectué en mode déconnecté.

Ce travail a été initié dans le cadre de l'action spécifique (AS) «Mobilité / Accès aux données» du CNRS, au sein de l'axe «cohérence» que j'ai animé. Il a donné lieu à deux stages de DEA. Les travaux de l'AS ont donné lieu à un article de synthèse/prospective en deux parties [BBB+01]. Une version plus courte de cet article a été publiée dans ACM Sigmod Records [BBB+04].

## 6.5 Partage de ressources scientifiques par médiation (2002-2005)

Les applications scientifiques de l'informatique ont depuis de nombreuses années accumulé de grande masses de données ainsi que des modèles et des programmes qui traitent et utilisent ces données à des fins d'analyse, de prévisions, de décision. Il s'est ainsi constitué au fil du temps un véritable réseau de ressources scientifiques, composé aussi bien de données observées (collectées sur le terrain par exemple) que de données produites par traitement informatique et des traitements eux-mêmes. L'objectif du projet PADOUE est l'exploitation globale de ce réseau. Elle nécessite la mutualisation de ces ressources qui sont par nature hétérogènes (fichiers, bases de données, programmes...) et dispersées sur les différents sites de production et de traitement. La mutualisation consiste à mettre les ressources à disposition sous un format homogène, mais aussi à offrir la possibilité de localiser ces ressources et de pouvoir définir et exécuter des processus mettant en œuvre plusieurs sources et traitements provenant de sites différents. Les résultats de ces processus doivent ensuite pouvoir être eux-mêmes intégrés au réseau de ressources.

Pour résoudre l'ensemble des problèmes posés, le projet PADOUE (ACI GRID) se base sur un système de médiation, Le Select, développé dans le projet Caravel à l'INRIA, qui offre une vue uniforme des ressources. L'intégration de nouvelles sources ou de nouveaux types de sources se fait par le développement d'adaptateurs spécifiques. Dans ce contexte, nos travaux ont porté sur la localisation des ressources à l'aide des métadonnées qui décrivent ces ressources. Afin d'équilibrer la charge du réseau, les métadonnées sont réparties sur différents sites et, pour assurer le passage à l'échelle, une architecture P2P (pair à pair) est utilisée [LDG05, LDG06].

Le projet s'est achevé en décembre 2005. Il a été jugé très satisfaisant par les experts du Ministère, qui soulignent son exemplarité en matière de collaboration entre informaticiens et domaines d'applications.

## 6.6 Gestion de données en Pair-à-pair (depuis 2006)

Le projet RESPIRE s'inscrit à la croisée des problématiques des projets PADOUE et Leg@net. Il s'agit de concevoir une infrastructure P2P pour le partage de données (bases de données, documents, etc.) et de service de gestion de données (stockage, évaluation de requêtes, vues, etc.). Les utilisateurs potentiels d'une telle infrastructure sont les communautés professionnelles (chercheurs, enseignants, architectes...), qui veulent pouvoir partager des ressources de diverses natures. Ces communautés étant par nature nombreuses à l'échelle de l'Internet et typiquement coopératives, une architecture P2P semble toute désignée. Ce projet, dont j'ai été le responsable, était financé par l'ANR dans le cadre des ARA (Action Recherche Amont) et regroupait le LIP6, l'IRISA l'INRIA, l'INT et le LINA (Nantes, ex-IRIN). Ce projet a démarré début 2006 et plusieurs collaborations entre différents

partenaires ont émergé sur des thématiques précises (réplication optimiste en P2P, regroupement de pairs, architecture, réplication en bases de données et en systèmes répartis). Si le projet s'est terminé fin 2008 (très bonne évaluation), mes axes de recherche sur la thématique continuent pour la plupart :

- Le premier axe consiste à adapter nos travaux sur la fraîcheur à des données XML, puisque ce modèle est un standard de fait pour les échanges dans les réseaux large-échelle. Ces travaux visent à résoudre plusieurs problèmes : trouver les mesures de fraîcheur pertinentes pour des documents XML et les algorithmes efficaces pour les calculer. Ceux-ci, de même que l'algorithme de routage, se basent sur un premier algorithme permettant de déterminer si deux expressions Xpath (l'une représentant une requête ou une mise à jour, l'autre une mise à jour) touchent des données en commun [GLL07].
- Le deuxième axe concerne le contrôle de concurrence pour les documents XML. Depuis peu, les applications utilisent XML non seulement pour rechercher l'information, mais aussi pour mettre à jour les documents, rendant nécessaire un mécanisme de contrôle de concurrence pour garantir la cohérence des accès. Contrairement à la plupart des approches qui étendent à XML les mécanismes de verrouillage classiques, nous avons étudié une approche optimiste (sans verrouillage), mieux adaptée aux environnements P2P. Nous nous basons sur l'analyse du code des transactions pour déterminer les conflits à l'aide de l'algorithme de détection cité au premier axe [BKGL08]
- Le troisième axe a été mené avec l'équipe Paris de l'IRISA, qui a mis au point et développé JuxMem, un service de partage de données en mémoire vive pour les grilles, basée sur l'environnement JXTA. Nos recherches visent à étendre JuxMem avec des fonctionnalités de SGBD, plus particulièrement l'indexation et le contrôle de concurrence, afin de gérer des bases de données de très grande taille en mémoire vive [AAB+07].
- Le quatrième axe est lié à la réplication et au relâchement de fraîcheur dans des systèmes à large échelle. Contrairement aux grappes de PC considérées dans Leg@net, l'Internet pose des problèmes de temps de réponse et de fiabilité du réseau, que nous devons intégrer dans les algorithmes de routage. De plus, les métadonnées nécessaires au routage doivent être réparties. Dans le cadre de la thèse d'I. Sarr, encadré par H. Naacke, nous proposons un répertoire réparti pour les métadonnées. Les gestionnaires de transaction du réseau se synchronisent pour garantir la cohérence globale des données [SNG08, SNG10c] et la tolérance aux pannes et aux déconnexions [SNG10a, SNG10b]. L'extension de ces travaux à des transactions s'exécutant sur plusieurs noeuds simultanément fera l'objet de la thèse de M. Thiongane que je devrais diriger à partir de l'an prochain (sous réserve d'acceptation de bourse).

## 6.7 Archivage du Web et préservation numérique (depuis 2008)

Le projet ANR CARTEC a pour objet l'archivage de sites Web dans le cadre du dépôt légal (partenaire INA). L'objectif global est l'analyse de la variation temporelle des ensembles de documents Web qui évoluent de façon anarchique afin de les transformer en corpus, c'est-à-dire en une ressource construite et exploitable. Dans le cadre de la thèse de Myriam Ben-Saad que j'encadre, nous cherchons à optimiser la complétude et la cohérence temporelle de l'archive et de son index. En effet, les sites pouvant modifier les pages à tout moment, il n'est pas envisageable de capter toutes les versions de toutes les pages. Il est donc important de capter les « bonnes versions » en prévoyant le comportement des sites à partir des captures déjà faites, mais aussi de capter des versions temporellement cohérentes afin de permettre une navigation cohérente dans l'archive. La thèse de Zeynep Pehlivan que je co-encadre a pour but l'utilisation de l'archive : recherche d'information, requêtes et navigation. Le problème est de tenir compte à la fois de la dimension temporelle de l'archive, de son incomplétude ainsi que du comportement propre aux utilisateurs d'archive (chercheurs en sciences humaines pour une grande part). [BGP09] et [BG10] présentent l'architecture envisagée, basée sur la segmentation des pages, sur un modèle d'importance des informations et des changements, ainsi

que sur un algorithme original de comparaison de versions de page.[PBG10] présente les détails de l'algorithme de comparaison et ses utilisations potentielles.

Plus récemment, nous nous intéressons à un problème connexe qui est celui de la maintenance des archives. Une fois constituée, une archive est amenée à évoluer pour continuer à être exploitable (changement de support, de format). Cette évolution doit être contrôlée afin de ne pas perdre en qualité. Notre approche est de concevoir des outils permettant de calculer la qualité de l'archive avant et après maintenance, ce qui permet d'en déduire la qualité de l'opération de maintenance. Ces travaux seront menés dans le cadre du projet européen SCAPE (FP7 - IP).

## 6.8 Perspectives

Mes perspectives de recherche à court terme sont décrites dans les deux sous-sections précédentes, correspondant aux thèses de Z. Pehlivan et M. Ben-Saad d'une part (archivage du Web), d'I. Sarr et M. Thiongane (gestion de données à large-échelle) d'autre part. A moyen terme, la problématique de l'archivage va engendrer le besoin de maintenir ces archives constituées. Les outils que nous développons dans le cadre du projet Cartec pourraient être exploités pour aider au contrôle de la qualité des opérations de maintenance (changement de format d'archive par exemple) dans le cadre du projet européen Scape. Plus précisément, nous souhaitons étendre notre algorithme de comparaison en y intégrant une composante de comparaison d'images (collaboration avec M. Cord du LIP6), afin de déterminer avec plus de précision si une page Web archivée (considérée à la fois comme une structure et comme une image) est détérioré ou pas par une opération de maintenance. Ces travaux feront l'objet pour le moins d'une thèse coencadrée avec M. Cord.

À terme, les recherches en bases de données vont suivre une évolution déjà entamée depuis des années, à savoir la modularisation des fonctionnalités de gestion de données. En effet, si certains travaux concernent l'intégration de nouvelles fonctionnalités au sein même des SGBD, ils correspondent à des évolutions coûteuses en terme d'intégration à des SGBD commerciaux. Ceux-ci sont des systèmes très complexes, dont toutes les fonctionnalités ne sont pas forcément nécessaires dans bien des applications actuelles. La tendance actuelle serait plutôt de repenser et d'adapter certains concepts et techniques de bases de données (ex. stockage, indexation, contrôle de concurrence et transactions) à des environnements et des applications qui en ont réellement besoin. L'avenir de mes recherches passe donc par l'élargissement de mon champ de compétences et de collaborations, aussi bien au niveau scientifique et technique qu'au niveau applicatif. Sur ce dernier point, mon ambition est d'aller vers des applications «réellement utiles» (aide aux pays en développement, aides aux handicapés, ...), afin de multiplier ma motivation de chercheur par celle de voir l'application réussir.

Sur la forme, j'envisage mon rôle de Professeur sur la base d'un équilibre entre responsabilités, animation et technique. Mon ambition est de former et mener une équipe de taille raisonnable, très réactive, capable de s'intégrer par des compétences spécifiques et reconnues à des projets variés, à l'intérieur de la structure qui m'accueillera comme aux niveaux national et international. Évidemment, ma responsabilité ne se limitera pas à une équipe, et j'assumerai toute fonction pour laquelle je me sentirai compétent et utile, comme je l'ai toujours fait depuis que je suis enseignant-chercheur.

## 7 Liste de publications

La plupart des publications sont accessibles à l'url [www-poleia.lip6.fr/~gancarsk/publications.html](http://www-poleia.lip6.fr/~gancarsk/publications.html)

### Revues internationales

[GJ01] S. Gancarski and G. Jomier. A framework for programming multiversion databases. *Data and Knowledge Engineering*, 36(1) :29 – 53, 2001.

- [BBB+04] G. Bernard, J. Ben-Othman, L. Bouganim, G. Canals, B. Defude, J. Ferrié, S. Gañçarski, R. Guerraoui, P. Molli, P. Pucheral, C. Roncancio, P. Serrano-Alvarado, and P. Valduriez. Mobile databases : a report on open issues and research directions. *ACM SIGMOD Records*, 33(2), 2004.
- [LG05] Cécile Le Pape and Stéphane Gañçarski. Data Quality Management in a Database Cluster with Lazy Replication. *Journal of Digital Information Management*, 3(2), 2005.
- [GNPV06] S. Gañçarski, H. Naacke, E. Pacitti, and P. Valduriez. The Leganet System : Freshness-aware Transaction Routing in a Database Cluster . *Information systems, Elsevier*, 32(2) : 320–343, 2007.

### Revues internationales en ligne

- [GLNRS06] S. Gañçarski, C. León, H. Naacke, M. Rukoz, and P. Santini. Integrity Constraint Checking in Distributed Nested Transactions over a Database Cluster. In *CLEI Electronic Journal*, 9(2), <http://www.clei.cl/cleiej/paper.php?id=121>, 2006.
- [LDG06] Nicolas Lumineau, Anne Doucet, and Stéphane Gañçarski. Thematic Schemas Building for Mediation-based P2P Architecture. In *Electronic Notes in Theoretical Computer Science, Elsevier*, 150(2) :21–36, <http://www.sciencedirect.com/science/journal/15710661>, 2006.

### Revues nationales

- [GJ95] S. Gañçarski and G. Jomier. Gestion des versions d’entité et de leurs contextes : Analyse et perspectives. *Ingénierie des Systèmes d’Information*, 3(6) :677 – 711, 1995.
- [DGJM96] A. Doucet, S. Gañçarski, G. Jomier, and S. Monties. Maintien de la cohérence dans une base de données multiversion (selected papers from bda’96). *Ingénierie des systèmes d’information*, 5(1) :91–120, 1997.
- [BBB+01] G. Bernard, J. Ben-Othman, L. Bouganim, G. Canals, B. Defude, J. Ferrié, S. Gañçarski, R. Guerraoui, P. Molli, P. Pucheral, C. Roncancio, P. Serrano-Alvarado, and P. Valduriez. Mobilité et bases de données : état de l’art et perspectives. *Technique et science informatiques, Hermès Science*, 22(3 et 4), 2003.
- [LG04] Cécile Le Pape and Stéphane Gañçarski. Fraîcheur et validité de données répliquées dans des environnements transactionnels. *Ingénierie des systèmes d’information*, 9(5-6), 2004.
- [SNG10a] I. Sarr, H. Naacke and S. Gañçarski. Routage Décentralisé de Transactions avec Gestion des Pannes dans un Réseau à Large Echelle. In *RSTI-ISI, special issue, selected papers BDA (Bases de données avancées) 08*, Hermès Science publ., 15(1), 2010.

### Chapitres de livre

- [CGJM01] W. Cellary, S. Gañçarski, G. Jomier, and M. Manouvrier. Les versions. In A. Doucet and G. Jomier , editor, *Bases de Données et Internet*, chapitre 7. Hermès Science, 2001.
- [DG01] A. Doucet and S. Gañçarski. Entrepôts de données et bases de données multidimensionnelles. In A. Doucet and G. Jomier , editor, *Bases de Données et Internet*, chapitre 12. Hermès Science, 2001.

### Conférences internationales

- [GJ94] S. Gañçarski and G. Jomier. Managing entity versions within their context : a formal approach. In *Database and Expert System Applications*, pages 400–409, LNCS 856, Athens (Greece), 1994.

- [GJZ95] S. Gañçarski, G. Jomier, and M. Zamfiroiu. A framework for the manipulation of a multiversion database. In *DEXA'95 Workshop*, pages 247–256, London (U.K.), 1995.
- [DGJM96a] A. Doucet, S. Gañçarski, G. Jomier, and S. Monties. Integrity constraints and versions. In *Workshop "Integrity in Databases"*, Dagstuhl (Germany), 1996.
- [DGJM96b] A. Doucet, S. Gañçarski, G. Jomier, and S. Monties. Integrity constraints in multiversion databases. In *British National Conference On Databases*, pages 57–73, LNCS 1094, Edinburgh (Ecosse), 1996.
- [GRW97] S. Gañçarski, J. Rykowski, and W. Wiczerzycki. On object and database versioning in distributed environment. In *Proc. Int'l Conf. ISCIS*, Antalya (Turkey), 1997.
- [DGLR00] A. Doucet, S. Gañçarski, C. León, and M. Rukoz. Nested transactions with integrity constraints. In *Transactions and Database Dynamics (selected papers)*, pages 130–149, LNCS 1773, Dagstuhl (Germany), 2000.
- [LBDGM00] A. Laurent, B. Bouchon-Meunier, A. Doucet, S. Gañçarski, and C. Marsala. Fuzzy data mining from multidimensional databases. In *Int. symposium on Computational Intelligence*, Springer-Verlag, Studies CI 54S, Kosice (Slovakia), 2000.
- [DGLR01] A. Doucet, S. Gañçarski, C. León, and M. Rukoz. Checking integrity constraints in multidatabase systems with nested transactions. In *International Conference on Cooperative Information Systems*, pages 316–328, LNCS 2172, Trento (Italy), 2001.
- [GNPV02] S. Gañçarski, H. Naacke, E. Pacitti, and P. Valduriez. Parallel Processing with Autonomous Databases in a Cluster System. In *International Conference on Cooperative Information Systems*, pages 410 – 428, LNCS 2519, Irvine (California) et aussi dans les actes de BDA'02, 2002.
- [GNV02] S. Gañçarski, H. Naacke, and P. Valduriez. Load balancing of autonomous applications and databases in a cluster system . In *Distributed Data and Structures*, pages 159 – 170, Carleton Scientific, Paris (France), 2002.
- [GLV03] S. Gañçarski, C. Le Pape, and P. Valduriez. Relaxing Freshness to Improve Load Balancing in a Cluster of Autonomous Replicated Databases. In *Distributed Data and Structures*, Thessaloniki (Greece), 2003.
- [LGV03] C. Le Pape, S. Gañçarski, and P. Valduriez. Trading Freshness for Performance in a Cluster of Replicated Databases. In *International Conference on Cooperative Information Systems*, pages 14– 15, Poster sessions, LNCS 2889 Catania, Sicily, (Italy), 2003.
- [LGV04] Cécile Le Pape, Stéphane Gañçarski, and Patrick Valduriez. Refresco : Improving query performance through freshness control in a database cluster. In *Int. Conf. On Cooperative Information Systems (CoopIS)*, pages 174–193, 2004.
- [GLN05] Stéphane Gañçarski, Cécile Le Pape, and Hubert Naacke. Fine-grained Refresh Strategies for Managing Replication in Database Clusters . In *VLDB'05 Workshop on Design, Implementation, and Deployment of Database Replication*, Trondheim (Norway), 2005.
- [LDG05] Nicolas Lumineau, Anne Doucet, and Stéphane Gañçarski. Thematic Schemas Building for Mediation-based P2P Architecture. In *InterDB Workshop on Database Interoperability*, Namur (Belgium), 2005 (aussi publié en revue en ligne, voir plus haut).
- [LG06] Cécile Le Pape and Stéphane Gañçarski. Replica Refresh Strategies in a Database Cluster. In *Proc. VECPAR (LNCS 4395)*, pages 679–691, Rio de Janeiro, (Brazil), also in BDA'05, 2006.
- [AAB+07] A. Almaksour, G. Antoniu, L. Bougé, L. Cudennec, and S. Gañçarski. Building a dbms on top of the juxmem grid data-sharing service. In *Proc. HiPerGRID Workshop*, Brasov, Romania, 15-15 September 2007. In conjunction with Parallel Architectures and Compilation Techniques 2007 (PACT2007).

- [GLL07] Stephane Gancarski, Cecile Le Pape, and Alda Lopes Gancarski. Freshness control of xml documents for query load balancing. In *Proc. of Xantec'07 (DEXA workshops)*, pages 35–39. IEEE Computer Society, 2007.
- [BKGL08] Djamel Berrabah, Stéphane Gañçarski, Sarah Kaddour Chikh, and Cecile Le Pape. Optimistic path-based concurrency control over xml documents. In *Proc. IEEE/ACM Fifth International Conference on Soft Computing as Transdisciplinary Science and Technology (CSTST'08)*, pages 390–396. ACM, 2008.
- [SNG08] Idrissa Sarr, Hubert Naacke, and Stéphane Gañçarski. Dtr : Distributed transaction routing in a large scale network. In *VECPAR (LNCS 5336)*, pages 521–531, 2008.
- [BGP09] Myriam Ben Saad, Stephane Gañçarski, and Zeynep Pehlivan. A novel web archiving approach based on visual pages analysis. In *Proc. of 9th International Web Archiving Workshop, collocated with ECDL*, Corfu, Greece, 2009.
- [SNG10b] Idrissa Sarr, Hubert Naacke, and Stephane Gañçarski. Failure-tolerant transaction routing at large scale. In *Proc. of International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA)*, Les Menuires, France, 2010.
- [SNG10c] Idrissa Sarr, Hubert Naacke, and Stephane Gañçarski. Transpeer : Adaptive distributed transaction monitoring for web2.0 applications. In *Proc. of ACM Symposium on Applied Computing : Track on Dependable and Adaptive Distributed Systems (SAC DADS)*, Sierre, Switzerland, 2010.
- [PBG10] Zeynep Pehlivan, Myriam Ben Saad, and Stéphane Gañçarski. Vi-diff : Understanding web pages changes. In *Proc. of 21st. DEXA Conference*, pages 1–15, Bilbao, Spain, 2010.
- [BG10] Myriam Ben Saad and Stephane Gañçarski. Using Visual Pages Analysis for Optimizing Web Archiving. In *Proc. of EDBT/ICDT Workshops, PhD workshop, Best Contribution Award*, Lausanne, Switzerland, 2010.

## Conférences nationales

- [GJ94] S. Gañçarski and G. Jomier. Un formalisme pour la gestion de versions d'entité dans leur contexte. In *BDA (french national conference on databases)*, Clermont-Ferrand (France), 1994.
- [GJ96] S. Gañçarski and G. Jomier. Vers un langage de manipulation pour bases de données multiversion. In *BDA (french national conference on databases)*, Cassis (France), 1996.
- [DFGJM96] A. Doucet, M.-C. Fauvet, S. Gañçarski, G. Jomier and S. Monties. Maintien de la Cohérence dans une base de Données Multiversion. In *BDA (french national conference on databases)*, Cassis (France), 1996.
- [GLR00] S. Gañçarski, C. León, and M. Rukoz. Vérification de contraintes dans les transactions imbriquées réparties. In *BDA (Bases de données avancées)*, pages 61–80, Blois, France, 2000.
- [LGM00] A. Laurent, S. Gañçarski, and C. Marsala. Coopération entre un système d'extraction de connaissances floues et un système de gestion de bases de données multidimensionnelles. In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 325–332, La Rochelle (France), 2000.
- [DGLR00] A. Doucet, S. Gañçarski, C. León, and M. Rukoz. Estrategias para verificar restricciones de integridad globales en multi base de datos con transacciones anidadas. In *Conferencia Latinoamericana de Informática*, Mexico City (Mexico), 2000.
- [DGLLR02] A. Doucet, S. Gañçarski, V. Leguizamo, C. León, and M. Rukoz. Un prototipo de manejador de transacciones anidadas con restricciones de integridad. In *Conferencia Latinoamericana de Informática*, Montevideo, (Uruguay), 2002.

- [GLNRS05] S. Gañarski, C. León, H. Naacke, M. Rukoz, and P. Santini. Verificación de Restricciones de Integridad en Transacciones Distribuidas sobre un cluster de Bases de Datos Relacionales. In *Conferencia Latinoamericana de Informática (CLEI)*, Cali, (Columbia), 2005.
- [SNG08b] I. Sarr, H. Naacke and S. Gañarski. Distributed Transaction Routing with Failure Management in a Large Scale Network. In *BDA (Bases de données avancées)*, Guilhaingranges, France, 2008.

## Divers

- [Gan03] S. Gañarski. *Introduction aux bases de données : notes de cours (2è édition)*. UPMC - UFR 922, 60 pages, 2003.