Rapport INRIA 1994 — Programme 1

Bases de Données

Projet VERSO

3 mai 1995

Projet VERSO

Bases de Données

Localisation : *Rocquencourt*

Mots-clés : base de données cartographiques (1), base multimédia (1), document électronique (1), génome (1), hypertexte (1), langage de base de données (1), modèle de base de données (1), optimisation de requête (1), SGBD à objets (1), théorie des bases de données (1).

1 Composition de l'équipe

Responsable scientifique

Serge Abiteboul, DR Inria

Responsable permanent

Stéphane Grumbach, CR Inria

Secrétaire

Danièle Moreau, AJA Inria

Conseillers scientifiques

Claude Delobel, Professeur, Université de Paris 11
Michel Scholl, Professeur, CNAM
Victor Vianu, Professeur, UC San Diego

Personnel Inria

Sophie Cluet, CR Inria

Chercheurs extérieurs

Bernd Amann, ATER puis Maître de Conférence, CNAM
Emmanuel Waller, ATER puis M.C. Paris 11

Chercheurs invités

Anuj Dawar, Professeur, U. College of Swansea, 1 mois
Tova Milo, Chercheur, U. Toronto, 6 semaines
Guido Moerkotte, Professeur, U. Karlsruhe, 1 mois
Jianwen Su, Professeur, UC Santa Barbara, 1 mois

Chercheurs post-doctorants

Gerd Hillebrand, boursier ERCIM, Brown U., 6 mois
Dimitri Theodoratos, boursier ERCIM, U. Athènes, Nov. 94
Jan Van den Bussche, boursier HCM, U. Anvers, Oct. 94

Chercheurs doctorants

Vassilis Christophides, Boursier MESR, CNAM
Laurent Herr, Boursier ENS
Zoé Lacroix, Boursier MESR ENSTA/LRI puis ATER LRI
Philippe Rigaux, Boursier Inria
Luc Ségoufin, Boursier ENS
Cassio Souza dos Santos, Boursier brésilien
Fariza Tahi, Boursier Inria, Action Génome

Stagiaires

Eric Angel, Magistère ENS, 2 mois
Karim Souak, DEA Orsay , 6 mois

2 Présentation du projet

L'objectif du projet est de développer des solutions novatrices aux problèmes posés par la gestion de bases de données (BD). Dans ce cadre, il nous paraît essentiel de poursuivre nos travaux sur trois niveaux: (i) l'investigation des principes fondamentaux du domaine, (ii) l'amélioration des systèmes de gestion de BD (SGBD) existants, et (iii) l'étude de nouvelles applications.

Les SGBD à objets ont des atouts majeurs comme la puissance de modélisation et la facilité de maintenance des programmes et des données. Cependant, pour être totalement satisfaisante, cette nouvelle technologie doit encore évoluer pour, notamment, offrir plus d'indépendance entre les niveaux physique et logique. C'est dans cette direction que se situe notre travail. A l'actif de l'équipe cette année, une extension du mécanisme de vues O_2Views, un travail sur la norme OQL, des techniques

d'optimisation de requêtes imbriquées et à nombreuses jointures, l'achèvement d'un interprète/optimiseur de requêtes, et enfin, des solutions concrètes à la migration des objets d'une classe à l'autre.

En amont de ces travaux, nos efforts se portent naturellement sur la compréhension des aspects fondamentaux des modèles de BD et de leurs liens avec le développement de nouveaux outils. Nos travaux ont permis de cerner l'expressivité et la complexité de nombreux langages pour BD (à objets), en prenant en compte de nouveaux types de données, la réflexivité et des aspects de la concurrence et du parallélisme.

Le cadre applicatif, en aval des deux premiers, est également privilégié et fournit une mine de nouveaux problèmes. Nous développons des modèles permettant le traitement des documents électroniques et des données hypertextes et géographiques. Également exigée par ce type de domaine (et par nombreux autres) est l'interopérabilité entres les applications bases de données et d'autres applications. Nous étudions l'utilisation des outils sophistiqués du monde base de données pour l'accès à des données contenues dans des systèmes de fichiers.

Notre travail complète naturellement les recherches réalisées dans le projet Rodin (Systèmes de Bases de Données) avec qui nous entretenons une collaboration étroite (séminaire et projets Esprit en commun). L'équipe Verso est aussi impliquée dans l'action Génome (compression et interrogation de séquences d'ADN, et prédiction de structures secondaires d'ARN).

Pour Verso, l'année 1994 a été marquée par l'amélioration d'anciens outils (O_2 Views, HyperPATH/O_2, interprète et optimiseur OSQL, etc.) et la définition d'une nouvelle norme OQL. Au niveau européen, Verso est fortement impliqué dans des Projets Esprit (Fide2, GoodStep) et dans le Réseau d'excellence Idomeneus.

Une activité autour des bases de données de documents s'est consolidée cette année et devrait être un des thèmes majeurs de l'équipe pour les années à venir. Des contacts avec des groupes académiques (en Europe et aux US) et des industriels (en France) ont été établis sur ce thème.

3 Action de recherche

3.1 Bases de données à objets

Notre travail sur les bases de données à objets se situe essentiellement autour de l'indépendance physique/logique, caractéristique fondamentale introduite par les systèmes relationnels. Notre recherche consiste, d'une part, à offrir des fonctionnalités manquantes aux systèmes actuels (vues, migration d'objets) et, de l'autre, à améliorer les fonctionnalités existantes (optimisation de requêtes, normalisation).

3.1.1 Mécanisme de vues

Participants : Serge Abiteboul, Claude Delobel, Cassio Souza dos Santos, Emmanuel Waller

Les vues apportent des facilités de restructuration et d'intégration de données, tout en permettant l'adaptation des structures de données aux besoins de différentes applications. Elles ont été proposées initialement pour les SGBD relationnels. Dans le contexte orienté-objet, outre la restructuration des données, elles permettent l'adaptation du comportement des objets.

Nous définissons dans [25] un langage de vues. Les primitives du langage permettent la redéfinition de l'interface des objets d'une base ainsi que la définition de nouveaux objets à partir d'objets existants. Une vue est définie comme étant un schéma virtuel dont on peut dériver des bases virtuelles. Une base virtuelle est l'image d'une base réelle à travers la vue. Les caractéristiques du monde objet sont, bien entendu, intégrées au mécanisme de vues et y jouent un rôle important.

Une formalisation partielle du mécanisme est présentée dans [26]. Une formalisation plus générale est en cours.

Le prototype O_2 Views constitue une base de validation de nos idées et a été implanté au-dessus du SGBD[1] O_2. Il met en œuvre un premier fragment du langage défini dans [25].

Par rapport à la première version datée de décembre 1993, le prototype actuel introduit l'héritage dans la vue et les classes paramétrées.

[1] O_2 est un SGBD à objets développé à l'INRIA dans le cadre du GIP Altaïr et commercialisé par O_2Technology.

Une extension du prototype O_2 Views permettant la migration des données d'une base relationnelle vers une base orientée objet a été développée par Karim Souak et est décrite dans [38].

Ce travail est réalisé dans le cadre du projet Esprit GoodStep [27].

3.1.2 Langages de requêtes et normes

Participants : Sophie Cluet, Claude Delobel

Dans le domaine des bases de données, les normes ont une place de plus en plus importante. La norme relationnelle SQL qui a vu le jour en 1986 a subi depuis cette date une première évolution avec la norme SQL2. En 1992, une nouvelle proposition a été faite pour introduire le paradigme objet dans le modèle relationnel. Cette proposition est en cours de discussion et sera probablement sujette à évolution et à révision. D'un autre côté, le groupe ODMG a proposé, dans le cadre de l'OMG, un document qui définit la norme des bases à objets et notamment le langage de requêtes OQL.

En collaboration avec la société O_2Technology, nous avons redéfini le langage OQL afin de le rendre compatible avec SQL(1). Cette nouvelle norme a été acceptée à l'unanimité par l'ODMG en octobre 1994.

Nous avons étudié les similitudes et les divergences qui existent entre la norme SQL3 et celle de l'ODMG en examinant le système de typage et les répercussions sur le langage de requêtes [24]. Nous mettons en évidence les différences d'approche entre ces deux normes. La norme SQL vise à créer un langage de programmation pour base de données incorporant les paradigmes objet et déductif, tandis que la norme ODMG propose l'intégration des fonctionnalités d'une base de données à objets dans des langages de programmation comme C++ ou Smalltalk.

3.1.3 Optimisation de requêtes

Participant : Sophie Cluet

La plupart des SGBD à objets propose maintenant un langage de requêtes déclaratif. Souvent ces langages n'ont de déclaratif que le nom et ne possèdent en général pas de réel optimiseur. De plus, l'évaluation des requêtes dépend en grande partie de leur formulation.

Cette année, nous avons travaillé autour de deux pôles: l'optimisation des requêtes imbriquées et l'optimisation des requêtes à nombreuses jointures.

Les requêtes imbriquées des langages orientés-objet sont essentielles pour plusieurs raisons. Elles permettent d'exprimer des conditions complexes, de construire ou d'accéder à des structures imbriquées. Leur optimisation restait cependant un problème ouvert. Nous proposons une classification des requêtes imbriquées dans le contexte orienté-objet ainsi que les techniques appropriées pour les optimiser [22]. L'optimisation consiste à *désimbriquer* les requêtes et est basée sur un système de réécriture algébrique.

Dans les systèmes à objets, chaque élément d'une expression de chemin représente une jointure implicite. De plus, les ensembles traversés par les expressions de chemin ont des tailles plus ou moins importantes et il est parfois intéressant d'envisager l'introduction de produits cartésiens. Nous étudions la complexité du problème de l'ordonnancement des jointures avec introduction possible de produits cartésiens et donnons un algorithme pour les requêtes en étoile dont la complexité est moindre que celle des méthodes jusqu'à présent utilisées [23].

En collaboration avec O_2Technology, nous avons implanté et testé un nouvel interprète du langage OQL offrant un meilleur niveau d'optimisation.

Le travail sur l'optimisation est partiellement financé par le projet Esprit Fide2.

3.1.4 Migration d'objets

Participants : Emmanuel Waller

Nous étudions un mécanisme qui supporte la migration d'objets d'une classe d'un SGBD à objets à une autre. Il est ainsi possible de modéliser la situation dans laquelle le même objet joue différents rôles au cours de sa vie. La migration d'un objet peut générer des conflits de type dus aux différentes contraintes de type imposées par les classes.

Nous présentons un mécanisme d'adaptation qui résout automatiquement ces conflits. Ce mécanisme combine reclassification des objets et modification des attributs. Nous étudions la complexité du problème,

et montrons en particulier que ce mécanisme est peu coûteux pour les schémas covariants.

3.2 Aspects fondamentaux des modèles bases de données

Bien comprendre les modèles de bases de données permet de les utiliser au mieux. Pour cette raison, la théorie a une place importante dans Verso.

3.2.1 Machines Relationnelles, Réflexivité et Parallélisme

Participants : Serge Abiteboul, Victor Vianu

Un modèle de calcul, appelé *machine relationnelle*, avait été introduit pour modéliser le style de calcul en bases de données consistant en un langage de requêtes (e.g., SQL) associé à un langage de programmation (e.g., C). Des résultats importants avaient permis de mieux cerner la puissance et les limites de ce modèle [8, 7].

Nous étendons cette machine en utilisant la génération dynamique de requêtes, une forme de réflexivité dans le modèle [17]. Nos résultats concernent la puissance des machines obtenues (*machines relationnelles réflexives*) sous certaines restrictions de ressources en temps et en espace, et suivant des limites sur le nombre de variables autorisées dans les formules logiques.

Il se trouve que ces machines procurent un modèle de calcul parallèle (formulé en termes logiques) analogue au modèle standard des circuits uniformes. Cette correspondance étroite avec les circuits qui n'existe pas dans le cas des machines relationnelles non-réflexives suggère que cette forme de réflexion est liée au parallélisme massif.

Ce travail théorique permet d'apporter des fondements logiques au style "moderne" (de plus en plus répandus) d'application bases de données: des postes de travail/clients utilisant des serveurs bases de données sur machines fortement parallèles. Cela devrait permettre de mieux comprendre le potentiel et les limites de tels systèmes.

3.2.2 Langages de Requêtes Fonctionnels

Participants : Serge Abiteboul, Gerd Hillebrand

De nombreux langages de requêtes fonctionnels ont été proposés, e.g., les langages relationnels et pour objets complexes, ou des propositions plus récentes autour de l'induction structurelle.

On utilise le λ-calcul simplement typé comme véhicule pour exprimer des requêtes imbriquées dans un langage de programmation [3, 30]. On montre que de nombreux langages déjà proposés peuvent être vus comme des sous-langages du λ-calcul simplement typé. Nous déterminons la complexité de tels langages en fonction de l'ordre des types.

Nous nous concentrons sur la complexité en espace pour l'évaluation de requêtes dans des langages de requêtes fonctionnels [16]. Nous étudions des optimisations qui permettent d'atteindre, avec les langages fonctionnels, l'efficacité des langages procéduraux notamment en utilisant du parallélisme.

3.2.3 Complexité des langages bases de données

Participants : Stéphane Grumbach, Victor Vianu

Les langages pour la manipulation d'objets complexes et leur complexité ont été abondamment étudiés et en particulier dans le Projet Verso. On montre comment définir des langages, dont la complexité est bornée polynomialement, pour des objets complexes [13]. La manipulation de multi-ensembles étend radicalement le pouvoir d'expression d'un langage de requêtes. Nous proposons dans [36] une famille d'algèbres pour multi-ensembles et étudions leur complexité. En présence de certains opérateurs, nous obtenons des algèbres dont la complexité varie de l'espace logarithmique à l'hyperexponentiel. Le gain de pouvoir d'expression est dû à un opérateur de multi-ensemble des parties. Nous étudions également dans [28] les ordres partiels sur les multi-ensembles (pomsets). Un tel ordre peut être représenté par un graphe acyclique orienté dont les nœuds peuvent avoir le même label. Ce type de données généralise tous les types classiques et en particulier l'ensemble et la liste. Nous définissons une algèbre dont le pouvoir d'expression est similaire à celui de la logique point-fixe avec compteurs.

Nous étudions dans [12] le pouvoir d'expression de la logique du premier ordre, de la logique inductive ainsi que de la logique infinitaire étendue avec certains quantificateurs généralisés. Nous montrons des résultats de probabilités asymptotiques d'énoncés de la logique du premier ordre avec des quantificateurs de Härtig (permettant d'énoncer la propriété qu'il y a autant de valeurs satisfaisant une propriété ϕ que de valeurs satisfaisant une propriété ψ). Ces résultats sont basés sur des preuves combinatoires et analytiques qui ont été réalisées en collaboration avec G. Fayolle (Projet Meval). Nous étendons des techniques de jeu d'Ehrenfeucht-Fraïssé aux langages avec compteurs, pour montrer des résultats de hiérarchie.

Nous proposons une extension du concept de définition implicite autorisant une forme de non-déterminisme au niveau du calcul [35]. Nous montrons en particulier que cette nouvelle définition capture précisément le pouvoir d'expression de la classe NP∩co-NP sur les structures finies. Ce type de résultats met en évidence le pouvoir d'expression des primitives non-déterministes.

3.2.4 Bases de données avec contraintes

Participant : Stéphane Grumbach

Les nouvelles applications de bases de données (par exemple l'espace) conduisent à des bases de données infinies, mais récursives, et donc admettant une représentation effective finie. De telles bases de données peuvent être représentées à l'aide de contraintes polynomiales sur les réels, par exemple. Nous étudions le pouvoir d'expression de la logique du premier ordre sur des classes restreintes de structures infinies, comme les ordres denses ou les corps ordonnés [29]. Nous définissons de nouvelles techniques pour vérifier la non-définissabilité d'une propriété au premier ordre [18]. Nous proposons des extensions des modèles avec objets structurés aux contraintes sur les ordres denses [37].

3.2.5 Bases de données actives

Participant : Laurent Herr

Les SGBD traditionnels sont passifs : ils n'exécutent des opérations que si l'utilisateur ou un programme le leur demande explicitement.

(Certains systèmes comportent un mécanisme de déclencheurs mais ces derniers sont difficiles à programmer, à maîtriser.) Or, de plus en plus d'applications, comme la bourse, la gestion de stocks et les réseaux informatiques, nécessitent des SGBD capables de réagir à des événements extérieurs ou à des modifications de la base de données. C'est pour répondre à cette demande qu'ont été conçus les SGBD actifs.

Des prototypes ont déjà été implantés. Mais ils se révèlent à l'usage peu adaptés aux applications nouvelles comme le multimédia ou la communication. Nous nous intéressons donc à une extension des langages actifs permettant de traiter ces applications. D'autre part, sur un plan plus théorique, nous étudions des modèles de SGBD actifs permettant de rendre compte de leur puissance d'expression et de leur complexité.

3.3 Nouvelles applications

Une partie importante du travail de l'équipe consiste à comprendre les problèmes posés par les nouvelles applications et à leur trouver des solutions élégantes et efficaces. Ce travail est doublement utile. Il permet la résolution de problèmes spécifiques, en l'occurrence l'écriture d'applications élaborées, mais il ouvre également la voie à la conception d'outils génériques pouvant être utiles à de nombreuses applications.

3.3.1 Bases de données géographiques

Participants : Michel Scholl, Philippe Rigaux

L'accent a porté cette année sur le choix d'index spatiaux pour optimiser les requêtes spatiales d'une base de données géographiques et sur la représentation et l'interrogation d'informations géographiques multi-échelle.

Nous avons évalué les performances de quatre index spatiaux implantés au-dessus du système O_2 [14]. L'objectif recherché était le choix de l'index spatial simple que doit comporter tout SGBD pour accélérer des requêtes spatiales fréquentes comme le pointé ou le fenêtrage. Parmi ces quatre index, figuraient deux structures d'arbres classiques : le *Quadtree* et le *R+tree*. Nous avons vérifié expérimentalement les avantages et inconvénients de ces deux types de structure: le quadtree est performant lorsque la charge est faible (index mémoire centrale). Il est très

facile à implanter. Le R+tree a de bonnes performances mais est extrêmement difficile à mettre en œuvre. Nous voulions également savoir si, pour des raisons de performances, ce type d'index doit être implanté dans les couches basses du SGBD ou s'il peut être développé en utilisant l'interface de programmation du SGBD. Il semble que la deuxième solution soit suffisante lorsque des primitives de regroupement sur disque sont disponibles dans le langage de développement du SGBD, et lorsque l'application comporte peu de mises à jour ce qui est très courant en Géographie.

Notre deuxième direction de recherche concerne la représentation multiple des objets géographiques. Suivant l'échelle d'une carte thématique, on peut vouloir montrer plus ou moins d'objets, plus ou moins de détails d'un objet. A un objet représenté à petite échelle par un point (par exemple, une ville), peut correspondre à plus grande échelle un ou plusieurs objets de géométrie différente (la ville devient un groupe de polygones). Un noyau d'interface géographique pour la représentation de tels objets a été réalisé en C++ et connecté à un prototype de SGBD géographique (GéO2) implanté par l'IGN au-dessus du système O_2 sur lequel une interrogation multi-échelle a été implantée [33].

Ces travaux ont été réalisés dans le cadre du projet Esprit Amusing, en collaboration avec le laboratoire Cedric du CNAM et avec l'aide de O2Technology et de l'IGN (mise à disposition du logiciel GeO2).

3.3.2 Bases de données et documents structurés

Participants : Serge Abiteboul, Vassilis Christophides, Sophie Cluet, Laurent Herr, Michel Scholl

Les documents électroniques représentent une large classe des données manipulées aujourd'hui : édition électronique, documentation, systèmes médicaux, etc. Malheureusement, ni les hypertextes ni les systèmes de recherche d'information (SRI) ne fournissent les fonctionnalités nécessaires pour la manipulation de documents: modélisation, langages de requêtes puissants et conviviaux, vues, concurrence d'accès, etc. C'est pourquoi l'intégration des techniques des SRI (e.g., recherches sur le contenu), des hypertextes (navigation) et des SGBD (e.g., recherches sur la structure) est primordiale. Un état de l'art sur la question a été fait [9] qui montre l'intérêt d'une telle approche.

Un modèle général a été défini pour représenter et manipuler des documents structurés au sein d'un SGBDOO [20]. A titre d'exemple, nous avons travaillé sur des documents SGML avec le SGBD O_2. Un langage de requêtes déclaratif est proposé qui permet d'interroger un document structuré par structure et par contenu. Le langage permet l'utilisation d'opérateurs textuels (à la SRI) et une navigation à travers la structure d'un document sans connaissance complète de cette structure.

Une traduction d'un fichier SGML en objets complexes O_2 est proposée ainsi qu'une extension du langage de requêtes OQL, basée sur l'introduction d'expressions de chemins dans le modèle. Ces chemins permettent de traverser structures et données complexes de manière uniforme. Finalement, l'application de cette approche aux hypertextes et notamment à l'interrogation de documents Hytime est présentée dans [21]. Bien que motivée par la manipulation de documents structurés, cette approche s'avère d'application beaucoup plus générale.

Une implantation avec le système O_2 a démarré par l'écriture d'un traducteur SGML/O_2. Ce traducteur est basé sur l'analyseur développé par la société Euroclid. Il génère pour chaque type (DTD) de document un schéma O_2 et pour chaque document des objets et valeurs correspondants dans la base de données. Cette traduction n'est pas encore satisfaisante et nous comptons améliorer ses performances, et offrir plus de flexibilité dans le choix du schéma bases de données cible dans des travaux futurs. L'étape suivante consiste à implanter dans le SGBD O_2 les fonctionnalités du langage proposé.

Un autre axe de nos travaux sur les documents, concerne l'interrogation et la mise à jour de documents stockés dans un système de fichiers. Dans une publication précédente, nous avions montré comment des données structurées stockées dans des fichiers pouvaient bénéficier des technologies développées dans le contexte des BD. Nous nous intéressons plus particulièrement au problème de la mise à jour des fichiers à l'aide d'un langage BD de haut niveau [15]. Nous définissons un cadre rendant ces mises à jour possibles et introduisons une technique d'exécution efficace. La technique est basée sur un mécanisme inverse d'une analyse grammaticale ("*unparsing*") et utilise également les techniques d'optimisation de requêtes précédemment introduites.

3.3.3 Bases de données et hypertextes

Participants : Bernd Amann, Michel Scholl

Le succès récent des systèmes et interfaces hypertextes (par exemple, WWW et Mosaic) s'explique par l'apparition de nouvelles applications qui demandent une organisation flexible et une exploitation simple de toutes sortes de documents. La création de réseaux d'information de taille de plus en plus importante pose de nouveaux problèmes qui sont principalement liés à la gestion et l'utilisation (navigation) des liens entre les documents. Lorsque le volume de données est trop important, la navigation devient fastidieuse et les risques de se perdre sont nombreux. De plus, l'absence de typage des données est un handicap certain à la compréhension de l'information et à son interrogation. La solution que nous préconisons repose sur l'intégration d'un logiciel hypertexte à un SGBD. Cette intégration (i) permet le stockage et l'accès aux documents et (ii) fournit la possibilité d'accéder aux documents par navigation et au moyen d'un langage de requêtes déclaratif.

L'an passé, nous avions défini le modèle hypertexte *Gram*. Nous avons ensuite travaillé à l'élaboration de mécanismes pour (i) la navigation assistée par des requêtes dans des graphes complexes au moyen de raccourcis, (ii) la mémorisation des chemins d'accès et (iii) la cohabitation de plusieurs bases de données et le passage de l'une à l'autre [2, 19].

Nous avons ensuite évalué notre proposition en utilisant une intégration du système hypermédia Multicard (développé par BULL S.A. et distribué par Euroclid) avec le SGBD O_2. L'extension de ces systèmes aux mécanismes d'interrogation et de navigation cités plus haut à abouti au prototype HyperO_2. Un aspect important dans ce prototype est son interface d'interrogation graphique qui permet aux utilisateurs/lecteurs de décrire des chemins dans les schémas des différentes applications [2].

4 Actions industrielles

Des liens étroits existent avec la société O_2Technology, le système O_2 nous servant de base principale d'expérimentation. Le mécanisme de vues O_2Views [25] et le système d'interrogation d'hypertexte HyperO_2 [2, 19] ont été implantés au-dessus du système O_2. L'interrogation de documents structurés en cours de développement utilise également ce

système. Par ailleurs, dans le cadre des travaux menés sur l'optimisation de requêtes, un interprète de requêtes OQL a été réalisé.

Des liens existent également avec la société Euroclid: la traduction de documents structurés utilise le traducteur SGML développé par cette société et le système d'interrogation d'hypertexte HyperO_2 utilise le logiciel Multicard.

5 Actions nationales et internationales

5.1 Actions nationales

5.1.1 PRC-GDR Bases de Données

L'équipe joue un rôle moteur dans le Programme de Recherches Coordonnées *Bases de Données* devenu cette année un GDR. S. Abiteboul, C. Delobel, et M. Scholl ont été impliqués dans le pilotage et l'administration du PRC/GDR. L'équipe participe à deux de ses pôles de recherches.

5.1.2 Génome

S. Grumbach et F. Tahi participent, dans le cadre de l'action Génome, aux recherches entreprises dans le thème *Informatique et Génome* [10, 11]. (Voir ce rapport d'activité, la section sur *Action Transversale Génome et Calcul*). M. Scholl est membre du conseil scientifique du Groupement d'Etudes et Recherches sur les Génomes (GREG), qui coordonne et finance les recherches françaises sur le génome.

Nous collaborons avec l'action Génome sur le thème d'une base de données génétiques.

5.1.3 Autres

A l'intérieur de l'Inria, des collaborations ont lieu avec les projets Rodin (séminaire commun, Projet Esprit Fide2, PRC/GDR, projet en cours avec deux universités israéliennes), Algo (recherche dans les génomes avec M. Régnier), Meval (G. Fayolle). Des liens étroits existent aussi avec le LIPN (N. Bidoit, C. Tollu), le LRI (Claude Delobel, Emmanuel Waller), le Cedric (M. Scholl, B. Amann), l'IMAG (équipe de M. Adiba) et l'ENST (V. Vianu).

Nous collaborons (M. Scholl) avec l'action Praxitèle sur le thème d'une base de données pour les transports.

5.2 Actions internationales

5.2.1 Projet Esprit BRA AMUSING

Ce projet terminé en Mai 1994 avait pour objet l'étude des problèmes rencontrés dans la gestion de données spatiales : interfaces utilisateur, modèles de données, langages de requêtes, index spatiaux, optimisation, développement de prototypes, etc. Les universités TU (Vienne), ETH (Zürich), Fern (Hagen), NTUA (Athènes), La Sapienza (Rome), le CNR (Rome) et la Société Algotech (Rome) y participaient. La France était représentée par l'IGN et le projet Verso. Le projet Verso (en collaboration avec le Cedric, CNAM) s'est plus particulièrement intéressé aux thèmes suivants : interfaces, index spatiaux et langages de requêtes.

5.2.2 Projet Esprit GOODSTEP

L'objectif du projet GoodStep est d'offrir un support bases de données pour le développement d'environnements de génie logiciel [27]. L'intitulé du projet est *General Object-Oriented Database for Software Engineering Processes*. Les universités de Francfort, Dortmund et Grenoble, les instituts de recherche Inria (équipe Verso) et Cefriel (Milan) y participent. Du côté industriel, on trouve O_2Technology, British Airways et Engineering (Padoue). L'objectif de Verso dans GoodStep est de fournir des outils avancés de modélisation. Nous avons conçu et implanté un mécanisme de vues dans le SGBD à objets O_2 [34, 25, 26]. Le prototype, disponible sous ftp, est actuellement utilisé notamment par des partenaires du projet dans le développement de leurs applications de génie logiciel.

5.2.3 Projet Esprit FIDE2

L'objectif du projet Fide2 est d'offrir un support pour la conception et l'implantation de systèmes dédiés à des applications bases de données de grande taille et de longue durée. L'intitulé exact du projet est : *Fully Integrated Data Environment 2*. Les universités de Glasgow, Hambourg, Pise et St Andrews, le centre de recherche IEI-CNR de Pise et l'Inria y

participent. L'Inria est représentée par les projets Rodin et Verso. Les intérêts de Verso dans Fide2 sont essentiellement axés sur la modélisation et l'optimisation de requêtes.

5.2.4 Réseau d'excellence IDOMENEUS

IDOMENEUS est un réseau européen d'excellence intitulé *Information and Data on Open MEdia for NEtworks of USers*. Son but principal est de coordonner et améliorer les efforts européens pour le développement des environnements d'information de demain capables de gérer et communiquer une classe importante d'informations sur un large ensemble de médias. Il se situe donc à la convergence des domaines des bases de données, du multimédia et de la recherche d'information. L'équipe est un des nœuds administratifs de ce réseau (responsable des échanges scientifiques entre les équipes du réseau).

5.2.5 Autres

En Amérique du Nord, des travaux en commun sont en cours avec l'Université de Toronto (A. Mendelzon, T. Milo) [15, 36, 28, 32], UC Santa Barbara (J. Su) [29, 37], et UC San Diego (V. Vianu). Nous avons eu la visite de Gerd Hillebrand (Brown University) comme post-doc sur financement ERCIM. S. Grumbach sera à Toronto en sabbatique à partir de Novembre 1994. L. Ségoufin devrait faire son service militaire à UCSD.

Pour l'Asie, l'équipe est impliquée dans le *Programme de Recherches Avancées Franco-Chinois*. S. Grumbach est responsable d'une collaboration avec l'Université FUDAN à Shanghai sur le thème des bases de données à objets.

Pour l'Europe, S. Grumbach collabore avec les universités de Swansea (A. Dawar) et d'Athènes (F. Afrati) [18]. S. Cluet collabore étroitement avec l'Université de Karlsruhe (G. Moerkotte) [22]. Nous avons la visite de Dimitri Theodoratos d'Athènes comme post-doc sur financement ERCIM. Des liens anciens avec l'Université d'Anvers se concrétisent par le post-doc de Jan van den Bussche dans Verso sur financement HCM. (Jan van den Bussche a obtenu le prix de la meilleure thèse de l'année en Belgique.)

Verso participe à deux projets EC-NSF: l'un lié au projet Esprit BRA Fide2 (avec S. Cluet) et l'autre intitulé *Non-déterminisme et bases de*

données (avec S. Grumbach). Verso est membre du réseau d'excellence Compulog et participe aussi à la création d'un réseau européen sur la théorie des modèles finis en cours de soumission. Enfin, Verso participe au groupe bases de données de l'ERCIM.

L'équipe a une longue tradition de collaboration avec des équipes israéliennes. Un contrat de collaboration (Arc en ciel) entre l'Inria (projets Rodin et Verso), l'Université Hébraïque (C. Beeri, Y. Sagiv) et l'Université de Tel Aviv (T. Milo) vient d'être soumis.

6 Diffusion des résultats

6.1 Enseignement

C. Delobel est professeur à l'Université de Paris 11. M. Scholl est professeur au CNAM-Paris et est co-responsable pour le CNAM du DEA SI (Paris 6, CNAM et ENST). V. Vianu est professeur à UCSD en sabbatique à l'ENST.

S. Abiteboul est maître de conférence à l'école polytechnique. B. Amann et E. Waller sont maîtres de conférence, respectivement au CNAM-Paris et à l'Université de Paris 11, depuis Octobre. Zoé Lacroix est ATER à l'Université Paris 11.

M. Scholl a présenté un cours sur la modélisation des données spatiales à la conférence EGIS/MARI à Paris (Mars 1994) et à la conférence BDA à Clermont-Ferrand (Août 1994). V. Christophides a présenté un cours sur la recherche documentaire par structure lors d'un cours INRIA, Aix en Provence, en Octobre [9].

Les cours suivants ont été assurés par divers membres de l'équipe.

SGBD relationnelles, CNAM-Paris, M. Scholl et B. Amann.

SGBD avancées, CNAM-Paris, M. Scholl et B. Amann; ENS-Ulm, S. Abiteboul; Ecole Polytechnique, S. Abiteboul.

SGBD à objets, DESS, Paris 11, S. Cluet et C. Souza dos Santos; DEA, Paris 11; C. Delobel; DEA, Paris 6, M. Scholl, V. Christophides.

Architecture matérielle et logicielle, Maîtrise, Paris 11. C. Delobel.

Architecture des gestionnaires d'objets, DESS, Université Paris 11. C. Delobel.

BD, MIAGE et nouvelle formation d'ingénieurs, Paris 11. C. Delobel; CNAM-Puteaux, V. Christophides.

Théories des bases de données, DEA BD commun Paris 1, 7 et 11, S. Grumbach.

6.2 Participation à des conférences et Colloques

L'équipe a eu de nombreuses publications dans des conférences internationales et des colloques (voir la bibliographie). De plus, certains membres de l'équipe ont participé à des comités de programmes. La liste en est donnée ci-dessous.

S. Abiteboul

- International Colloquium on Automatas, Languages and Programming, Jerusalem, Israel (1994) **PC Chairman**, et Hongrie (1995)
- International Conference on European Data Base Technology Cambridge, UK (1994)
- International Symposium on Logical Foundations of Computer Science, St. Petersburg, Russia (1994)
- 11èmes Journées Bases de Données Avancées, Clermont-Ferrand (1994) **Président des Journées**
- International Conference on Data Engineering, Taiwan (1995)
- ACM SIGACT-SIGMOD-SIGART Conference on Principles of Database Systems, San Jose (1995) **PC Chairman**

S. Cluet

- International Conference on Very Large Databases (VLDB), Santiago, Chili (1994)
- Sixth International Workshop on Persistent Object Systems (POS), Tarascon, France (1994)
- International Conference on Very Large Databases (VLDB), Zurich, Suisse (1995)
- 11èmes Journées Bases de Données Avancées (BDA), Nancy, France (1995)

C. Delobel

- International Conference on Very Large Databases (VLDB), Zurich, Suisse (1995) **Tutorials chairman**.

S. Grumbach

- 5th International Conference on Database Theory, Prague, République Tchèque (1995)

M. Scholl

- International Workshop on Geographical Information Systems, IGIS'94, Ascona, Switzerland (1994)
- Conférence AFCET INFORSID, France, (1994)
- Conférence Interfaces des mondes réels et virtuels, Montpellier (1994)

Victor Vianu

- ACM SIGACT-SIGMOD-SIGART Conference on Principles of Database Systems, Minneapolis (1994) **PC Chairman**
- International Conference on Very Large Databases (VLDB), Santiago, Chili (1994)

6.3 Organisation de colloques et de cours

Serge Abiteboul est coorganisateur des *Journées Sémantique et Bases de Données*, Lieusaint (Oct. 1994).

6.4 Diffusion de Produits

Le prototype O_2 Views est disponible par ftp. (Il faut disposer de O_2 pour l'utiliser.) Ce package est constitué d'un compilateur du langage de vues, d'une interface graphique pour le gestionaire de vues et d'un manuel d'utilisateur [34]. L'implantation du prototype est décrite en détail dans [26].

6.5 Autres

Victor Vianu a été présentateur invité au *Logic Colloquium 94, the Association for Symbolic Logic (ASL)* en Juillet 1994 à Clermont-Ferrand. V. Vianu et S. Grumbach ont présenté des communications au *Workshop on Logic and Computational Complexity* à Indianapolis en Octobre

1994. S. Grumbach a aussi présenté un exposé au workshop sur les quantificateurs généralisés, Helsinki, Février 94. Bernd Amann a participé à une table ronde dans le cadre des journées AFCET Jeunes (Nov 94).

S. Cluet a été invitée à présenter ses travaux par l'Université de Toronto; Stéphane Grumbach a été invité dans les universités de Toronto et de Santa Barbara (UCSB); Victor Vianu par celles de Pennsylvanie, d'Aachen, de Muenster, de Vienne et d'Oslo; et Serge Abiteboul par celles de Glasgow, Toronto et Pennsylvanie.

7 Publications

Livres et monographies

[1] S. Abiteboul, R. Hull, V. Vianu, *Foundations of Databases*, Addison-Wesley, 1994.

Thèses

[2] B. Amann, *Interrogation d'Hypertextes*, thèse de doctorat, Conservatoire National des Arts et Métiers, Paris, France, février 1994.

[3] G. Hillebrand, *Finite Model Theory in the Simply Typed Lambda Calculus*, thèse de doctorat, Brown University, 1994.

Articles et chapitres de livre

[4] S. Abiteboul, C. Beeri, «On the power of languages for complex values», *Very Large Data Bases Journal*, Nouvelle version. A paraître.

[5] S. Abiteboul, P. Kanellakis, «Object Identity as a Query Language Primitive», *Journal of the Association for Computing Machinary*, Nouvelle version. A paraître.

[6] S. Abiteboul, P. Kanellakis, E. Waller, «Method Schemas», *Journal of Computer Science and Systems*, Nouvelle version. A paraître.

[7] S. Abiteboul, M. Vardi, V. Vianu, «Computing with Infinitary Logic», *Theoretical Computer Science*, Papier invité. A paraître.

[8] S. Abiteboul, V. Vianu, «Computing with First-Order Logic», *Journal of Computer and System Sciences*, Papier invité. A paraître.

[9] V. Christophides, «Recherche Documentaire par Structure: une approche comparative entre SRI et SGBD», *in: Cours INRIA: Le Traitement Électronique de Document*, A. Editions (réd.), Aix en Provence, Octobre 1994.

[10] S. Grumbach, F. Tahi, «A New Chalenge for Compression Algorithms: Genetic Sequences», *Journal of Information Processing and Management 30 (6)*, 1994, Special Issue on Compression.

[11] S. Grumbach, F. Tahi, «Compression et compréhension des séquences de nucléotides», *Technique et Science Informatiques*, 1995, A paraître.

[12] S. Grumbach, C. Tollu, «On the Expressive Power of Counting», *Theoretical Computer Science*, 1994, Papier invité. A paraître. Aussi Rapport de recherche Inria, 2330, Sept 1994.

[13] S. Grumbach, V. Vianu, «Tractable Query Languages for Complex Object Databases», *Journal of Computer and System Sciences*, 1994, Papier invité. A paraître.

[14] J. Peloux, G. Reynal, M. Scholl, «Evaluation of spatial indices implemented with the DBMS O_2», *Ingénierie des systèmes d'information (ISI)*, 1994, A paraitre.

Communications à des congrès, colloques, etc.

[15] S. Abiteboul, S. Cluet, T. Milo, «More on Updating the File», *in : 10èmes Journées Bases de Données Avancées*, Clermont-Ferrand, France, Août 1994.

[16] S. Abiteboul, G. Hillebrand, «Space Usage in Functional Query Languages», *in : Int. Conf. on Database Theory*, Prague, République Tchèque, Janvier 1995.

[17] S. Abiteboul, C. Papadimitriou, V. Vianu, «The Power of the Reflective Relational Machine», *in : IEEE Symposium on Logic in Computer Science*, 1994.

[18] F. Afrati, S. Cosmadakis, S. Grumbach, G. Kuper, «Expressiveness of linear vs. polynomial constraints in database query languages», *in : Second Workshop on the Principles and Practice of Constraint Programming*, 1994.

[19] B. Amann, A. Rizk, M. Scholl, «Querying Typed Hypertexts in Multicard/O_2», *in : ACM European Conference on Hypermedia Technology*, p. 198–205, Edimbourg, Ecosse, Septembre 1994.

[20] V. Christophides, S. Abiteboul, S. Cluet, M. Scholl, «From Structured Documents to Novel Query Facilities», *in : Proc. ACM Sigmod*, Minneapolis, Minnesota, 1994.

[21] V. Christophides, A. Rizk, «Querying Structured Documents with Hypertext Links», *in : ACM European Conference on Hypermedia Technology*, Edimbourg, Ecosse, Septembre 1994.

[22] S. Cluet, G. Moerkotte, «Classification and Optimization of Nested queries in Object Bases», *in : 10èmes Journées Bases de Données Avancées*, Clermont-Ferrand, France, Août 1994.

[23] S. Cluet, G. Moerkotte, «On the Complexity of Generating Optimal Left-Deep Processing Trees with Cross-Products», *in : Int. Conf. on Database Theory*, Prague, République Tchèque, Janvier 1995.

[24] C. Delobel, «Convergence and/or divergence between object-oriented query languages and relational: ODMG and SQL3», *in : BIWIT*, Biarritz, France, février 1994.

[25] C. S. dos Santos, C. Delobel, S. Abiteboul, «Virtual Schemas and Bases», *in : International Conference on Extending Data Base Technology*, Cambridge, March 1994.

[26] C. S. dos Santos, «Design and Implementation of an Object-Oriented View Mechanism», *in : 10èmes Journées Bases de Données Avancées*, Clermont-Ferrand, France, Août 1994.

[27] R. Z. et al., «The GoodStep Project: General Object-Oriented Database for Software Engineering Processes», *in : Proc. Asian Pacific Software Engineering Conference*, Tokyo, Japon, 1994.

[28] S. Grumbach, T. Milo, «An Algebra for Pomsets», *in : Int. Conf. on Database Theory*, Prague, République Tchèque, Janvier 1995.

[29] S. Grumbach, J. Su, «Finitely representable Databases», *in : 13th ACM Symp. on Principles of Database Systems*, Minneapolis, Minnesota, Mai 1994. Papier invité *Journal of Computer and System Sciences.*

[30] P. Kanellakis, G. Hillebrand, H. Mairson, «An Analysis of the Core-ML Language: Expressive Power and Type Reconstruction», *in : Proceedings of the Twenty-First International Colloquium on Automata, Languages and Programming*, Jerusalem, Israel, 1994.

[31] S. Lifschitz, V. Vianu, «A probabilistic view of datalog parallelization», *in : Int. Conf. on Database Theory*, Prague, République Tchèque, Janvier 1995.

[32] A. Mendelzon, T. Milo, E. Waller, «Object migration», *in : Proceedings of ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems*, Minneapolis, 1994.

[33] P. Rigaux, M. Scholl, «Multiple Representation Modelling and Querying», *in : IGIS'94*, S. Verlag (réd.), *LNCS*, Ascona, Switzerland, février 1994.

Rapports de recherche et publications internes

[34] C. S. dos Santos, *The O_2 Views User's Manual - Upgrade of version 1*, juillet 1994.

[35] S. Grumbach, Z. Lacroix, «Non-deterministic Implicit Definitions», *rapport de recherche*, Verso - Inria, 1995.

[36] S. Grumbach, T. Milo, «Towards Tractable Algebras for Bags», *rapport de recherche*, Verso - Inria, 1994, Papier invité *Journal of Computer and System Sciences.*

[37] S. Grumbach, J. Su, «Dense Order Constraint Databases», *rapport de recherche*, Verso - Inria, 1995.

[38] K. Souak, *Migration d'une base relationnelle vers une base orientée objet : une approche par les vues*, Mémoire, Université Paris I/Paris XI, 1994.

8 Abstract

The goal of the Verso group is to develop novel solutions to problems of database management. This involves the improvement of existing system functionalities. Fundamental studies are also conducted to lay the foundations of such systems. Finally, specific applications are considered in order to validate the results and suggest new improvements.

Object-oriented database systems are becoming popular in the industrial world. However, to be really competitive, this new technology must be improved and, in particular, it must provide more independence between the physical and logical layers. This is a main theme of the group. A view mechanism on top of the O_2 system has been developped, new query optimization techniques introduced and concrete solutions for object migration proposed.

The theoretical part of the research done at Verso is concerned with understanding the fundamental aspects of database models and their links with the development of new tools. An important direction of research is the study of the complexity and expressiveness of database languages, taking into account aspects such as new data types, reflexivity, concurrency and parallelism.

The specific applications that we are considering are hypertext, electronic document and geographical database applications. For instance, we are interested in high level access languages and optimization in such contexts.

Keywords: database model, database theory, object-oriented database, multimedia base, geographic database, deductive database, hypertext, electronic documents

Some scientific cooperations: Participation in Esprit BRA Fide2 project, Esprit R&D GoodStep project, in Esprit Idomeneus Network of Excellence, in the Ercim group for databases; participation in two joint EC-NSF contracts; Involvement in Franco-Chinese advanced research programs, collaboration with the Hebrew and Tel Aviv Universities.

Table des matières