

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

—«0»—

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR

—«0»—

Université des Sciences et de la Technologie

« HOUARI BOUMEDIENE »

—«0»—

INSTITUT D'INFORMATIQUE

—0—

Projet de Fin d'Etudes

pour l'obtention du diplôme d'ingénieur

d'état en informatique

Option : SOFTWARE

THEME

OPTIMISATION DES REQUETES
DANS UN SYSTEME DE GESTION
DE BASE DE DONNEES
GENERALISEES

Présenté par Mlles Lamia FRIHA et Fatiha TAHI

JURY :

Mme G. TRIKI-YAMANI

Mlle N. BENSAOU

Mlle F. BETTOUCHE

Promotrices :

Mlle AZROU

Mlle F. BENDJENAH

Mlle BOUMALHA

Centre de Recherche sur l'Information Scientifique et Technique

(C E R I S T)

R E M E R C I E M E N T S

Nos profonds remerciements s'adressent à Monsieur M. BENHAMADI, Directeur du C.E.R.I.S.T, pour nous avoir accueillis au sein du centre et avoir mis à notre disposition les moyens matériels nécessaires à la réalisation de notre projet.

Nos remerciements s'adressent également à Madame G. TRIKI-YAMANI pour nous avoir honoré en acceptant la présidence du jury.

Nous remercions chaleureusement Mademoiselle N. BENSOU, membre de la sous-commission d'examen, qui nous a guidé tout au long de notre projet. Ses conseils et ses critiques nous ont été d'un apport bénéfique. Nous lui exprimons ici notre profonde gratitude pour avoir bien voulu persévérer avec nous en acceptant d'être membre du jury.'

Nos sincères remerciements vont également à Mademoiselle F. BETTOUCHE, membre de la sous-commission d'examen, qui a eu la gentillesse de lire et corriger ce mémoire. Nous lui exprimons notre reconnaissance pour son soutien et la sympathie dont elle a fait preuve à notre égard. Sa participation au jury nous honore.

Nous remercions vivement Mademoiselle F. AZROU, attachée de recherche au C.E.R.I.S.T, qui a bien voulu nous proposer ce sujet, pour les marques de confiance qu'elle nous a témoigné et pour ses critiques constructives.

Nous adressons également nos sincères remerciements à Mesdemoiselles F. BENDJENAH et H. BOUMALHA, ingénieurs au C.E.R.I.S.T, pour leur aide précieuse tant morale que professionnelle.

Nous exprimons notre profonde reconnaissance à Monsieur C. FRASSON, Professeur à l'université de Montréal, spécialisé dans les performances des bases de données, pour avoir accepté de lire et critiquer nos rapports lors de son passage à Alger. Nous le remercions vivement pour ses éclaircissements, ses conseils, ses encouragements et surtout pour l'intérêt qu'il a porté à notre travail.

Nos chaleureux remerciements et notre profonde gratitude vont à Monsieur R. KAL, Maître conférencier à Genève, pour son soutien en matière de documentation, sa disponibilité à notre égard et l'aide morale et matérielle qu'il nous a fourni tout au long de notre projet.

Enfin, pour tout le personnel du C.E.R.I.S.T, pour tous ceux qui de près ou de loin nous ont soutenu moralement, nous disons un grand merci.

R E S U M E

Dans le système de gestion de bases de données généralisées (SGBDG) qui est en cours de réalisation au C.E.R.I.S.T, l'indépendance entre la description logique des données et leur implantation physique sur les supports de stockage cache, en réalité, la grande complexité des différents algorithmes implantés dans ce système. L'exécution de ces algorithmes est réalisée au détriment du temps de réponse, ce qui diminue les performances du système.

La présente étude a pour but d'apporter, pour le SGBDG, un remède au problème de la dégradation du temps de réponse des requêtes utilisateurs. La solution consiste à améliorer la formulation et l'exécution de la requête en tenant compte des possibilités offertes par le système.

L'optimisation des requêtes réalisée porte sur deux axes: D'une part, elle procède à la réduction des tailles des résultats intermédiaires en recherchant l'ordonnement optimal des opérateurs apparaissant dans la requête et en choisissant un algorithme efficace pour exécuter l'opération de jointure lorsque celle-ci se présente.

Et d'autre part, elle opère une sélection appropriée d'un chemin d'accès aux données en estimant le coût d'accès pour chaque chemin candidat et en ne gardant que celui qui minimise le temps des entrées-sorties et de traitement dans l'unité centrale.

Cette première version de l'optimiseur MORDGEN (Module d'optimisation des Requêtes sur les Données Généralisées), traite les requêtes portant sur des tables de n-uplets. L'un des aspects évolutifs du système concerne la prise en compte de nouvelles méthodes d'accès et d'un ensemble d'opérateurs propres aux données complexes. La structure modulaire de MORDGEN permet l'adjonction aisée d'autres traitements spécifiques à ce type de données.

Mots clés

Système de gestion de bases de données - Optimisation -

- Chemin d'accès -

SOMMAIRE

INTRODUCTION .

PARTIE I

CHAPITRE I : ENVIRONNEMENT DE L'ETUDE .

I-1 / Concepts utilisés .

- I-1-1 / Notion d'entité .
- I-1-2 / Notion de surrogate .
- I-1-3 / Notion de fonction attribut .
- I-1-4 / Notion de type d'entités .
 - 1-1-4-1 / Type d'entités associatif .
 - 1-1-4-2 / Type d'entités caractéristique .
 - 1-1-4-3 / Type d'entités Kernel (Noyau) .
 - 1-1-4-4 / Type d'entités Kernel complexe .
 - 1-1-4-5 / Type d'entités agrégat .
 - 1-1-4-6 / Type d'entités contenu .
- I-1-5 / Notion de type de données DOCUMENT .
- I-1-6 / Notion de généralisation .
 - 1-1-6-1 / Spécialisation .
 - 1-1-6-2 / Union.

I-2 / Représentation des données sur les supports de stockage .

- I-2-1 / Structure de stockage des données simples .
- I-2-2 / Structure de stockage des données complexes .

CHAPITRE II : PRESENTATION DES TECHNIQUES D'OPTIMISATION .

II-1 / Stratégie d'optimisation dans SYSTEM-R .

- II-1-1 / Optimisation algébrique des requêtes .
- II-1-2 / Sélection du meilleur chemin d'accès aux données .
 - II-1-2-1 / Traitement des requêtes mono-relations
 - II-1-2-2 / Traitement des requêtes multi-relations
- II-1-3 / Conclusion .

II-2 / Stratégie d'optimisation dans INGRES .

- II-2-1 / Décomposition des requêtes multi-variables .
 - II-2-1-1 / Détachement d'une sous requête .
 - II-2-1-2 / Substitution de tuples .
 - II-2-1-3 / Exemple .
- II-2-2 / Traitement des requêtes mono-variables .
- II-2-3 / Conclusion .

11-3 / Technique des indices de jointure .

11-3-1 / Principe de la technique .

11-3-1-1 / Fonction de partitionnement .

11-3-1-2 / Fonction de duplication .

11-3-2 / Utilité des indices de jointure .

11-3-3 / Conclusion .

11-4 / Techniques d'optimisation combinatoires .

11-4-1 / Terminologie utilisée .

11-4-2 / Critère d'arrêt .

11-4-3 / Description des techniques d'optimisation combinatoires.

11-4-3-1 / Perturbation Walk (PW) .

11-4-3-2 / Quasi-random Sampling (QS) .

11-4-3-3 / Local Optimization (LO) .

11-4-3-4 / Simulated Annealing (SA) .

11-4-4 / Conclusion .

PARTIE II

CHAPITRE III : OPTIMISATION ALGEBRIQUE DES REQUETES DANS LE SGBDG.

III-1/ Présentation des opérateurs du SGBDG.

III-1-1/ Les opérateurs ensemblistes .

III-1-1-1/ Union.

III-1-1-2/ Intersection.

III-1-1-3/ Différence.

III-1-2/ Les opérateurs relationnels .

III-1-2-1/ Selection.

III-1-2-2/ Projection.

III-1-2-3/ Jointure.

III-1-2-4/ Projection sur types d'entités.

III-2/ Description de l'optimiseur algébrique du SGBDG.

III-2-1/ Réduction des sous-arbres homogènes.

III-2-2/ Distribution de la selection.

III-2-3/ Distribution de la projection sur types d'entités.

III-2-4/ Distribution de la projection.

III-2-5/ Elimination des tables vides et idempotentes.

CHAPITRE IV : SELECTION DES CHEMINS D'ACCES DANS LE SGBDG.

IV-1/ Les méthodes d'accès du SGBDG .

IV-1-1/ Le balayage séquentiel .

IV-1-2/ Le balayage d'index .

IV-2/ Description du mécanisme de sélection des chemins d'accès.

IV-2-1/ Les paramètres d'optimisation du mécanisme de sélection.

IV-2-2/ Estimation du coût d'accès à un type d'entités.

IV-3/ Le processus d'optimisation .

IV-3-1/ Optimisation d'une requête portant sur un type d'entités.

IV-3-2/ Optimisation d'une requête de jointure de N types d'entités.

CONCLUSION .

BIBLIOGRAPHIE.

ANNEXE: structures de données et algorithmes.