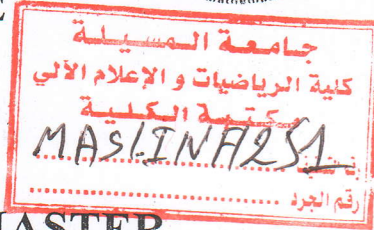


UNIVERSITE MOHAMED BOUDIAF - M'SILA
FACULTE DES MATHÉMATIQUES ET
DE L'INFORMATIQUE



DÉPARTEMENT D'INFORMATIQUE



MEMOIRE de fin d'étude
Présenté pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Systèmes d'Informations Avancés

Par: BARKA Ahlem

SUJET

**Méthodes bio-inspirées pour la fragmentation d'entrepôt
de données : Etude et évaluation**

Soutenu publiquement le : 01 / 06/2016 devant le jury composé de :

Dr. H. Belouadeh
Dr. C. Lamiche
A. Guerna

Université de M'sila
Université de M'sila
Université de M'sila

Président
Rapporteur
Examineur

Promotion : 2015/2016

Table des matières

-Liste des figures.....	i
-Liste des Tables.....	iii
-Liste des Algorithmes	iv
-Introduction générale	1
CHAPITRE I : Entrepôts de données : Concepts Fondamentaux.	
1. Introduction.....	2
2. L'entrepôt de donnée (ED).....	2
2.1.Définition.....	2
2.2.Caractéristiques.....	2
2.3.Les types de données.....	3
2.4.L'architecture d'entrepôt de données.....	4
2.4.1. Sources de données.....	5
2.4.2. Processus Extract, Transform, Load (ETL).....	5
2.4.3. Les magasins.....	5
2.4.4. Les applications clients.....	6
3. Les modèles des entrepôts de données.....	7
3.1.La modélisation conceptuelle.....	7
3.1.1. Concept de fait, de dimension et de hiérarchie.....	7
3.1.2. Modèles en étoile, en flocon et en constellation.....	7
3.1.2.1. Modèle en étoile.....	7
3.1.2.2. Modèle en flocons.....	8
3.1.2.3. Modèle en constellation.....	9
3.2.Modélisation logique.....	9
3.2.1. les modèles multidimensionnels.....	9
3.2.1.1. Modèle R-OLAP.....	10
3.2.1.2. Modèle M-OLAP.....	11
3.2.1.3. Model H-OLAP.....	11
3.3.La conception physique des entrepôts de données.....	11
4. Choix des techniques d'optimisation.....	11
4.1.Les techniques redondantes.....	12
4.1.1. Les indexe.....	12

4.1.1.1. Index Bitmap.....	12
4.1.1.2. Index B-Arbre.....	13
4.1.1.3. Les index de jointure binaires.....	14
4.1.2. La Fragmentation verticale.....	14
4.1.3. Les vues matérialisés.....	15
4.2. Les techniques non redondantes.....	16
4.2.1. La fragmentation horizontale.....	16
4.2.2. La fragmentation mixte.....	17

5. Conclusion.....

CHAPITRE II : La fragmentation horizontale dans les entrepôts de données

1. Introduction.....	18
2. La fragmentation horizontale FH.....	18
2.1. Fragmentation primaire.....	18
2.2. Fragmentation dérivée.....	19
2.3. Les avantages de la fragmentation horizontale.....	20
3. Méthodologie de fragmentation horizontale dans les entrepôts de données.....	20
4. Le Procédure de FH.....	21
4.1. Module de gestion des requêtes	22
4.2. Règles de correction de la fragmentation horizontale.....	22
5. Le problème de sélection d'un schéma de fragmentation.....	22
5.1. Travaux de recherche sur le problème de fragmentation.....	23
5.1.1. Travaux de DATTA et AL. (1999).....	23
5.1.2. Travaux de Wu et Buchmaan (1997).....	23
5.1.3. Travaux de NOAMAN & BARKER (1999).....	23
5.1.4. Travaux de Bellatreche (2000).....	23
6. Algorithmes de sélection d'un schéma de FH.....	23
6.1. Algorithmes basées sur les prédicats.....	24
6.2. Algorithmes basés sur l'affinité.....	24
6.3. Algorithmes à base de modèle de coût.....	27
7. Conclusion.....	28

CHAPITRE III : Métaheuristiques et fragmentation d'entrepôt de données.

1. Introduction.....	29
2. Description du problème.....	29

3. Les Algorithmes génétiques.....	30
3.1. Définition des algorithmes génétique.....	30
3.2. Principe de base.....	30
3.2.1. Le codage des individus.....	32
3.2.2. L'évaluation d'un individu.....	32
3.2.3. Les opérateurs génétiques.....	33
3.2.3.1. L'opérateur de sélection.....	33
3.2.3.2. L'opérateur de croisement ou crossover.....	34
3.2.3.3. L'opérateur de mutation.....	35
4. Les recuits simulés.....	35
5. Application de GA au Problème de sélection de schéma de FH.....	36
5.1. Processus génétique de la fragmentation horizontale.....	36
5.1.1. Codage adopté.....	36
5.1.2. Génération de la population initiale.....	37
5.1.3. Sélection des individus.....	37
5.1.4. Croisements de individus.....	37
5.1.5. Mutation.....	38
5.1.6. Fonction d'évaluation : <i>fitness</i>	38
5.1.7. Critère d'arrêt.....	38
6. L'application de recuit simulé RS sur la solution de GA.....	40
7. Conclusion.....	41

CHAPITRE IV : Réalisation et expérimentation

1. Introduction.....	42
2. Les outils de développement.....	42
2.1 Microsoft Visual Studio.....	42
2.1.1 Microsoft Visual Studio 2012.....	42
2.1.2 Langage de programmation C#	43
2.2 Oracle.....	43
2.2.1 Oracle Développer.....	44
3. Implémentation	44
3.1 L'entrepôt de données (Entrepôt de banc d'essai Benchmark).....	44
3.1.1. Types de requêtes prises en compte	45
3.1.2. Vecteur de sélectivité	46
3.2. Implémentation de solution	46

3.2.1. Procédure de solution	46
3.2.2. L'architecture de l'application.....	47
3.2.3. les cas d'utilisation de l'interface	47
4. Présentation graphique	48
5. Résultats obtenus.....	49
5.1 Analyse des résultats.....	51
6. Conclusion.....	51
-Conclusion général.....	52
- Bibliographique	53
-Annexes.....	58

Introduction générale

De nos jours, les entrepôts décisionnels d'entreprise constituent des ressources critiques. Avec l'évolution des exigences métiers, ils doivent gérer des volumes et des vitesses de données sans cesse grandissants. Les entrepôts sont construits autour d'une table de faits, souvent très volumineuse, qui peut atteindre plusieurs téraoctets, entourée de plusieurs tables de dimensions qui contiennent les axes d'analyse relativement aux indicateurs ou mesures contenus dans la table des faits. Ces derniers sont accédés par des requêtes décisionnelles complexes caractérisées par de multiples opérations de sélection, de jointure et d'agrégation. Pour optimiser ces requêtes, l'administrateur est amené à effectuer une tâche importante dans la conception physique, durant cette phase l'administrateur choisit un ensemble de techniques d'optimisation à sélectionner.

La fragmentation horizontale est une structure non redondante d'optimisation de requêtes OLAP. Dans le cadre de ce mémoire, nous formalisons d'abord le problème de sélection d'un schéma de fragmentation pour un entrepôt de données relationnel comme un problème d'optimisation. Nous proposons ensuite une méthode hybride combinant un algorithme génétique **AG** et un algorithme de recuit simulé **RS**.

Ce mémoire est organisé de la manière suivante : Chapitre 01 est consacré aux définitions des concepts fondamentaux d'entrepôt de données, aussi aux techniques d'optimisations des requêtes. Le deuxième chapitre présente d'une façon détaillée la technique d'optimisation *Fragmentation Horizontale FH* dans les entrepôts de données. Et concerne le troisième chapitre nous avons spécifié le problème générale qui on vas le traité, et après l'application des algorithmes génétique et des recuit simulé sur ce problème, Nous présentons aussi dans ce chapitre le modèle de coût utilisé. Dans le quatrième chapitre, nous avons présenté nos résultats de validation sur un banc d'essai APB-1 et un ensemble de 8 requêtes. Enfin, une conclusion rappelle la problématique étudiée ainsi que les résultats obtenus. Cette conclusion nous a permis de lister quelques perspectives de prolongement de nos travaux.

Conclusion générale

Dans le cadre de ce travail de master, nous avons vu et traité la complexité et la difficulté du problème de sélection de la technique de fragmentation horizontale dans un entrepôt de données. Au cœur de ce mémoire, nous avons présenté les différentes étapes de l'adaptation et la réalisation de notre application, nous avons commencé par la présentation générale des entrepôts de données, définition, caractéristique et l'architecture, ainsi les techniques d'optimisation des entrepôts de données. Nous avons présenté ensuite en détail la technique d'optimisation FH ou *fragmentation horizontale*, et après la détermination de problème de fragmentation horizontal « Sélection d'un schéma de fragmentation horizontale ». Et pour résoudre ce problème nous avons présenté en générale deux métaheuristiques: les algorithmes génétiques **GA** et recuis simulé **RS** et après nous avons adapté les **GA** pour résoudre le problème de sélection et l'algorithme **RS** pour améliorer les résultats d'algorithme génétique.

L'analyse des résultats présenté montre clairement que l'algorithme **GA** amélioré le cout de schéma de fragmentation (nombre Entré/Sortie) en fonction de nombre de générations, et aussi les résultats présentent l'amélioration de solution de **GA** avec l'hybridation '**GA+SR**'.

Comme perspectives à ce travail, Il est préférable d'exploiter la solution retournée par **GA** et **GA+RS** pour fragmenter l'entrepôt de données et ensuite l'exécution de la charge de requêtes sur les fragments trouvés pour réduire le temps d'exécution de ces requêtes.

into automate physical database design. Sigmod (SIGMOD 94), June 2004.

[7] B. Frédéric et T. Olivier, Construction graphique d'entrepôts et de magasins de données, IRIT (Institut de Recherche en Informatique de Toulouse), équipe SIG, Université Toulouse III.

[8] C.Favre, F. Benayeh, O.Boukhalid, J. Darmoni, G. Gavril, N. Harbi, N. Kabbaj et S. Loudcher, Les entrepôts de données pour les métiers du pas 1, Université de Lyon ERIC Lyon 2 et ERIC Lyon 1.

[9] Celec, Innovation dans la recherche et la production de statistiques: importance des nouvelles technologies. Flunla 2000, 27-28 Octobre 2000.

[10] E. Benitez, C. Collet, et M. Garcia, Entrepôts de données : caractéristiques et problématique. Revue TSI, 2003, 2001.

[11] E. ZIYATI, Optimisation de requêtes OLAP en Entrepôts de Données Approche basée sur la fragmentation générique. Thèse de Maitre de Recherche UNIVERSITÉ MOHAMMED VADDAL, 08 Mai 2010.

Bibliographie

- [1] A. DAÂS, mémoire magister :Optimisation des requêtes dans le datawarehouse, ECOLE DOCTORALE D'INFORMATIQUE (STIC) Option : Ingénierie des Systèmes Informatiques
- [2] A. Gherboudj, THESE doctorat : Méthodes de résolution de problèmes difficiles académiques, Université de Constantine2, Faculté des Technologies de l'information et de la Communication, Département d'Informatique Fondamentale et ses Applications, 2013
- [3] A. HASSAN, Modélisation des Bases de Données Multidimensionnelles : Analyse par Fonctions d'Agrégation Multiples, thèse de doctorat, UNIVERSITÉ TOULOUSE 1 CAPITOLE, 1décembre 2014.
- [4] A. KERKAD, THESE doctorat : L'interaction au service de l'optimisation à grande échelle des entrepôts de données relationnels,Ecole Doctorale : Sciences et Ingénierie pour l'Information, Mathématiques Secteur de Recherche : INFORMATIQUE ET APPLICATIONS, 2015.
- [5] A. Noamanetk. Barker, « A Horizontal Fragmentation Algorithm for the Fact Relation in a Distributed Data Warehouse », in the 8th International Conference on Information and Knowledge Management, November, 1999.
- [6] A. Sanjay, V. R. Narasayya, and B. Yang, Integrating vertical and horizontal partitioning into automated physical database design. Sigmod (SIGMOD'04), June 2004.
- [7] B. Frédéric et T. Olivier,Construction graphique d'entrepôts et de magasins de données, IRIT (Institut de Recherche en Informatique de Toulouse), équipe SIG, Université Toulouse III.
- [8] C.Favre, F. Bentayeb, O.Boussaid, J. Darmont,G. Gavin, N Harbi, N Kabachi et S Loudcher, Les entrepôts de données pour les nuls. . . ou pas !, Université de Lyon ERIC Lyon 2 et ERIC Lyon 1
- [9] Ceiec, Innovation dans la fourniture et la production de statistiques: importance des nouvelles technologies, Finland, 20 et 21 janvier 2000
- [10] E. Benitez, C. Collet, et M. Adiba. Entrepôts de données : caractéristiques et problématique. Revue TSI, 20(2), 2001.
- [11] E. ZIYATI, Optimisation de requêtes OLAP en Entrepôts de Données Approche basée sur la fragmentation génétique, thèse de doctorat, UNIVERSITÉ MOHAMMED VAGDAL, 08 Mai 2010.

Bibliographie

- [12] E. Ziyati, L. Bellatreche et K. boukhalfa « La contribution des structures d'optimisation non redondantes dans la conception physique des entrepôts de données » (ISPS'2007 algiers, Alger).
- [13] Hadj MAHBOUBI, «Optimisation de la performance des entrepôts de données XML par fragmentation et répartition », thèse de doctorat, Université Lumière Lyon 2, 08/12/2008
- [14] J.-F. Desnos, Entrepôt de données – Introduction, 2004
- [15] K. Boukhalfa, De la conception physique aux outils d'administration et de tuning des entrepôts de données, Mémoire de doctorat, Ecole Doctorale : Sciences pour l'Ingénieur et Aéronautique, 2009.
- [16] K. Boukhalfa et L. Bellatreche, Combinaison des algorithmes génétique et de recuit simulé pour la conception physique des entrepôts de données, Université de Laghouat, Algérie et LISI/ENSMA France, 2006.
- [17] K. Boukhalfa, L. Bellatreche et B. Ziani, Index de Jointure Binaires: Stratégies de Sélection & Étude de Performances, USTHB - Bp 32 El-Alia Alger, Algérie boukhalk@ensma.fr , LISI/ENSMA - Université de Poitiers , Futuroscope 86960 France bellatreche@ensma.fr ,Université de Laghouat – Algérie bziani@mail.lagh-univ.dz.
- [18] K. Boukhalfa, L. Bellatreche, et Z. Alimazighi, OptAssist : outil d'assistance pour l'optimisation des entrepôts de données relationnels, USTHB Université - Algiers – Algeria, LISI/ENSMA - Poitiers Université – France ,2006.
- [19] K. Ghali, Méthodologie de conception système à base de plateformes reconfigurables et programmables, Mars 2005.
- [20] L. Bellatreche, Techniques d'optimisation des requêtes dans les data warehouses, LISI/ENSMA Téléport2, Avenue Clément Ader ,86960 Futuroscope France, bellatreche@ensma.fr
- [21] L. Bellatreche. Utilisation des vues matérialisées, des index et de la fragmentation dans la conception logique et physique des entrepôts de données". Phd. thesis, Université de Clermont-Ferrand II, Décembre 2000.
- [22] L. Bellatreche et K. Boukhalfa, La fragmentation dans les entrepôts de données : une approche basée sur les algorithmes génétiques, Revue des Nouvelles Technologies de l'Information (EDA'2005), Juin, 2005.
- [23] L. BELLATRECHE, K. BOUKHALFA, et S. Caffiau « ParAdmin: Un Outil d'Assistance à l'Administration et Tuning d'un Entrepôt de Données », rapport de recherche, LISI/ENSMA - Université de Poitiers – Futuroscope.

Bibliographie

- [24] L. BELLATRECHE, K. BOUKHALFA et P. RICHARD, Rapport de recherche N° 01 - 2008 : Fragmentation Primaire et Dérivée: Étude de Complexité, Algorithmes de Sélection et Validation sous ORACLE10g, Université de Poitiers, France 2008.
- [25] L. Bellatreche, K. Boukhalfa et P. Richard. Referential horizontal partitioning selection problem in datawarehouses: Hardness study and selection algorithms. International Journal of Data Warehousing and Mining, 2004
- [26] L. Nathalie, Y. Pouillet., 2008. Revue du Droit des Technologies de l'Information - n° 30/2008, Doctrine Entrepôts de données et vie privée.
- [27] M. BARR, Mémoire de magistère : Approche dirigée par les fourmis pour la fragmentation horizontale des entrepôts de données relationnels, Ecole nationale Supérieure d'Informatique, 2009
- [28] M. BARR et L. BELLATRECHE, Approche dirigée par les fourmis pour la fragmentation horizontale dans les entrepôts de données relationnels, Revue « Nature & Technologie ». n° 06/Janvier 2012.
- [29] M. Bouzeghoub, ENTREPÔTS DE DONNEES ARCHITECTURE ET FONCTIONNALITES, université de Versailles SAINT-QUETIN-EN-YVELINES, Laboratoire PRISM <http://www.prism.uvsq.fr>.
- [30] M. Merouani et E. Abdelouarit , Les index pour les entrepôts de données : comparaison entre index arbre-B et Bitmap, Faculté poly disciplinaire de Tétouan et Faculté des sciences de Tétouan, <https://hal.archives-ouvertes.fr/hal-00842579> , 9 Jul 2013
- [31] M. T. Özsu and P. Valduriez. Principles of Distributed Database Systems ,1999 .
- [32] M. Teisseire, H. Alatriza Salas et F. Bouillot, Déroulement Du cours: Entrepôt De données et l'Analyse en ligne, (EPSI l'école d'ingénieurs informatique)
- [33] M. VAN CANEGHEM, « Le voyageur de commerce Algorithme "branch and bound", Algorithme Glouton, Méthode de recherche locale », support de cours, Département d'Informatique de la Faculté des Sciences de Luminy Université de la Méditerranée LIM : Laboratoire d'Informatique de Marseille, Décembre 2002.
- [34] O. Teste, Modélisation et manipulation d'entrepôts de données complexes et historisées, thèse doctorat, UNIVERSITE PAUL SABATIER DE TOULOUSE (SCIENCES), 18 décembre 2000.

Bibliographie

- [35] P. Frédérique Et T.Odile, A propos d'un entrepôt de données universitaire : modélisation des acteurs et méta données, LORIA, Campus Scientifique; Frederique.Peguiron@loria.fr , Odile.Thiery@loria.fr
- [36] R. BOUCHAKRI, thèse doctorat : Conception physique statique et dynamique des entrepôts de données, Écoles Doctorales, Sciences et Technologies de l'Information et de la Communication (STIC, ESI) Sciences et Ingénierie pour l'Information, Mathématiques (S2IM, ISAE-ENSMA), 17 Septembre 2015
- [37] R. BOUCHAKRI, Mémoire Pour l'obtention du diplôme de Magistère : Une approche dirigée par la classification des attributs pour fragmenter et indexer des entrepôts de données, (ESI) (Oued Semar, Alger), 2009
- [38] R. Bouchakri et L. Bellatreche, Sélection Incrémentale d'un Schéma de Fragmentation Horizontale d'un Entrepôt de Données Relationnel, Université de Poitiers France, Ecole nationale Supérieure d'Informatique (ESI) Alger.
- [39] R. Duvigneau, Introduction aux Méthodes d'Optimisation sans Gradient pour l'Optimisation et le Contrôle en Mécanique des Fluides, Ecole de printemps OCET Optimisation et Contrôle des Ecoulements et des Transferts, 12 - 17 Mars 2006.
- [40] R. Godin, C. Desrosiers, LOG660 - Bases de données de haute performance : Les entrepôts de données et l'analyse de données, Département de génie logiciel et des TI, Hiver 2011.
- [41] R. Moussa, Systèmes de Gestion de Bases de Données Réparties & Mécanismes de Répartition avec Oracle, Ecole Supérieure de Technologie et d'Informatique à Carthage, Année Universitaire 2005-2006.
- [42] S. Amédée et R. Francois-Gérard, ALGORITHMES GENETIQUES, TE de fin d'année Tutorat de Mr Philippe Audebaud, 21/06/2004.
- [43] S. BENKRID, THESE doctorat : Le déploiement, une phase à part entière dans le cycle de vie des entrepôts de données : application aux plateformes parallèles, École nationale Supérieure d'Informatique
- [44] S. SAICHI, Mémoire Pour l'obtention du diplôme de Magistère : Optimisation de requêtes dans les entrepôts de données, UNIVERSITE D'ORAN ES-SENIA, FACULTE DES SCIENCES DEPARTEMENT D'INFORMATIQUE, 27 juin 2009

- [45] S Silhadi-Hacid et M Tarafi Du DataWarehouse au WebHouse : le croisement des réseaux et des bases de données
- [46] S. Ramkumar, Algorithmes génétiques, 26 avril 2007
- [47] T. Martin, Microsoft Visual basic 2012-2013: mon eFormation,(page 25).
- [48] T. Olivier, Modélisation et manipulation d'entrepôts de données complexes et historisées, Décembre 2000.
- [49] T. Valléeet M.Yıldızoglu, Présentation des algorithmes génétiques et de leurs applications en économie, Université de Nantes et Université Montesquieu Bordeaux IV, 7 septembre 2001.
- [50] V G. Tourreau, C#, Publie par Pearson Education France, 2010.
- [51] W. Inmon. H. Building the Data Warehouse. John Wiley & Sons, 1996.
- [52] http://igm.univ-mlv.fr/~drXPOSE2013tleroux_genetic_algorithmindex.html.
- [53] <http://www.egilia.com/formation-oracle-10g-11g>, consulté le : 10/05/2016
- [54] <http://computertrainingcenters.com/top-10-features-of-visual-studio-2012>, consulté le : 5/05/2016
- [55] Wikipidia, <http://www.Wikipidia.com/ Visual Studio>,consulté le : 10/05/2016

```

Si on avec une probabilité  $\frac{1}{\text{NouvelCoût} + \text{Coût}}$ 
EtatCourant := NouveauEtat ;
Coût := NouveauCoût ;
Fin si ;
Si (Coût < CoûtMin) alors
EtatMin = EtatCourant ;
CoûtMin = Coût ;
Fin si ;
Jusqu'à l'obtention d'un équilibre // Nombre d'itérations
3.2 Réduire la température ;
Jusqu'à gel du système ; // Température = 0
4. Retourner EtatMin
Fin ;

```

الملخص :

إكتسبت مستودعات البيانات مكانة كبيرة في وقتنا الراهن لأهميتها الواضحة في اتخاذ القرارات حيث اعتبر الباحثون تطويرها وتسهيل الوصول إلى المعلومات أولوية كبيرة. وتعتبر تقنية التجزئة الأفقية من أبرز واحسن التقنيات في تحسين مستودع. في هذا العمل المخصص لمذكرة الماستر قمنا بتطبيق طريقة تهجين بين الخوارزميات الجينية و خوارزمية التخمير المحاكى لحل مشكل تحديد مسار التقسيم الأفقي لمستودع البيانات حيث استخدمنا مستودع البانات بنشمارك باعتماد وكلغة برمجة اخترنا سي شارب.

الكلمات المفتاحية: مستودعات البانات ، الخوارزميات الجينية، التخمير المحاكى، بنشمارك، الأوراكل و السي شارب

Résumé :

Les entrepôts de données occupent une place importante dans nos jours, grâce à son grand rôle dans la prise de décision, les chercheurs ont considéré que son développement est en priorité. Parmi les meilleures techniques d'optimisation des entrepôts de données c'est la fragmentation horizontale.

Dans notre travail, nous avons appliqué une méthode hybride combinant un algorithme génétique et un algorithme de recuit simulé pour résoudre le problème de Sélection d'schéma FH où en utilisant le banc d'essai Benchmark APB-1 en comptant sur Oracle, et comme langage de programmation on a choisi C #.

Mots clés : Entrepôt de données, Algorithmes génétiques, recuit simulé, Benchmark, Oracle et C#.

Abstract:

The data warehouse gained a big standing in our current time for its clear importance in making decision, where the researchers considered its development and facilitation of the arrival to the information big priority. The horizontal division Technique considers of the most prominent and best techniques in improving warehouse. In this specified work for master dissertation, we propose a hybrid method combining a genetic algorithm and a simulated annealing algorithm to solve the problem of determining horizontal partitioning of data warehouse where we have used data warehouse Benchmark adoption and chose C # as a programming language.

Keywords: Data warehouses, the genetic algorithms, simulated annealing, Benchmark, Oracle and C#.