

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٨٠ ماي ١٩٤٥ بسيدي بلعباس
École Supérieure en Informatique -08 Mai 1945- Sidi Bel Abbès



MÉMOIRE DE FIN D'ÉTUDE

Pour l'obtention du diplôme d'Ingénieur d'Etat en Informatique

Filière : Informatique

Spécialité : Systèmes d'Information et Web (SIW)

Thème

The Study of The Integration of The Variation of The Join Order
In a Materialized View Selection System Based on AI

Présenté par : M. MERZOUK ilyes reda

Soutenu le : **06 Juillet 2023** Devant le jury composé de

Dr. MALKI Abdelhamid	President
Dr. AWAD Samir	Examinateur
Dr. KECHAR Mohamed	Encadrant
Pr. BELLATRECHE Ladjel	Co-Encadrant

Année Universitaire : 2022/2023

Acknowledgments

First and foremost, I express my deepest gratitude to Allah for providing me with guidance, patience, and strength throughout this journey. His unwavering support has been instrumental in my perseverance and determination to complete this task successfully.

I would also like to extend my heartfelt appreciation to my parents, whose unwavering guidance and constant encouragement have shaped my path in life. Their invaluable advice and unwavering belief in my abilities have been the driving force behind my achievements. I am forever grateful for their sacrifices and dedication to my education and personal growth. It is with great pleasure and deep emotion that I dedicate the fruits of my efforts to them. I strive to fulfill their unconditional pride in me, knowing that their unwavering love and support have been my pillars of strength.

I extend my heartfelt gratitude and sincere thanks to my supervisor, **Mr. Kechar Mohamed**, for his unwavering trust and support throughout this endeavor. His guidance, expertise, and invaluable insights have played a pivotal role in the successful completion of this work.

I would like to express my deepest gratitude to **Mr. Ladjel Bellatreche**, professor at ISAE-ENSMA for generously dedicating his time and expertise to guide me during my internship. His pedagogical and scientific insights have been instrumental in shaping the direction of my work.

I would like to extend my sincere gratitude to all the members of the ESI-SBA community, including my fellow students, staff, and professors, for their invaluable contributions to my education and overall experience. Their dedication and collective effort have played an integral role in shaping the person I have become today, and I will forever cherish the memories and connections forged during my time at ESI-SBA.

Abstract

With the exponential growth of data in various domains, including cloud computing and other large-scale data environments, the efficient management and retrieval of information have become crucial challenges. This thesis explores the significance of optimizing queries and databases in such environments, focusing on the utilization of advanced techniques, including materialized views, to enhance performance, scalability, and overall system efficiency. It also addresses the key challenges that need to be addressed for effective optimization in large-scale data environments.

These environments often span multiple servers, data centers, or even geographical regions. Coordinating query execution across distributed resources while minimizing communication overhead and ensuring data consistency is a complex task. Optimizing query performance in such distributed environments requires the exploration of distributed query processing techniques, data partitioning strategies, and workload balancing mechanisms.

In conclusion, this thesis emphasizes the importance of addressing key challenges in optimizing queries and databases for large-scale data environments. By tackling issues related to scalability, distributed query processing, view maintenance, view selection, and integration with advanced optimization techniques, this research aims to contribute to the development of efficient and effective solutions for optimizing query performance and database management in the era of large-scale data.

Key Words: Query performance, Materialized views, Database management, View selection, Query evaluation.

Résumé

Avec la croissance exponentielle des données dans divers domaines, y compris l'informatique en nuage et d'autres environnements de données à grande échelle, la gestion efficace et la récupération des informations sont devenues des défis cruciaux. Cette thèse explore l'importance de l'optimisation des requêtes et des bases de données dans de tels environnements, en se concentrant sur l'utilisation de techniques avancées, y compris les vues matérialisées, pour améliorer les performances, l'évolutivité et l'efficacité globale du système. Elle aborde également les principaux défis à relever pour une optimisation efficace dans les environnements de données à grande échelle.

Ces environnements s'étendent souvent sur plusieurs serveurs, centres de données ou même régions géographiques. Coordonner l'exécution des requêtes sur des ressources distribuées tout en minimisant les coûts de communication et en garantissant la cohérence des données est une tâche complexe. L'optimisation des performances des requêtes dans de tels environnements distribués nécessite l'exploration de techniques de traitement des requêtes distribuées, de stratégies de partitionnement des données et de mécanismes d'équilibrage de la charge de travail.

En conclusion, cette thèse souligne l'importance de relever les principaux défis liés à l'optimisation des requêtes et des bases de données pour les environnements de données à grande échelle. En abordant les questions liées à l'évolutivité, au traitement distribué des requêtes, à la maintenance des vues, à la sélection des vues et à l'intégration avec des techniques d'optimisation avancées, cette recherche vise à contribuer au développement de solutions efficaces et efficientes pour optimiser les performances des requêtes et la gestion des bases de données à l'ère des données à grande échelle.

Mots clés : Performance des requêtes, vues matérialisées, gestion des bases de données, sélection des vues, évaluation des requêtes.

مُلخَص

مع النمو الهائل للبيانات في مختلف المجالات ، بما في ذلك الحوسبة السحابية وغيرها من بيئات البيانات واسعة النطاق ، أصبحت الإدارة الفعالة واسترجاع المعلومات تحديات حاسمة. تستكشف هذه الأطروحة أهمية تحسين الاستعلامات وقواعد البيانات في مثل هذه البيئات ، مع التركيز على استخدام التقنيات المتقدمة ، بما في ذلك الآراء المادية ، لتحسين الأداء وقابلية التوسع وكفاءة النظام بشكل عام. كما أنه يعالج التحديات الرئيسية التي يجب معالجتها من أجل التحسين الفعال في بيئات البيانات واسعة النطاق.

غالباً ما تمتد هذه البيئات عبر خوادم متعددة أو مراكز بيانات أو حتى مناطق جغرافية. يعد تنسيق تنفيذ الاستعلام عبر الموارد الموزعة مع تقليل عبء الاتصال وضمان تناسق البيانات مهمة معقدة. يتطلب تحسين أداء الاستعلام في مثل هذه البيئات الموزعة استكشاف تقنيات معالجة الاستعلام الموزعة واستراتيجيات تقسيم البيانات وآليات موازنة عبء العمل.

في الختام ، تؤكد هذه الأطروحة على أهمية معالجة التحديات الرئيسية في تحسين الاستعلامات وقواعد البيانات لبيئات البيانات واسعة النطاق. من خلال معالجة القضايا المتعلقة بقابلية التوسع ومعالجة الاستعلامات الموزعة وصيانة العرض واختيار العرض والتكامل مع تقنيات التحسين المتقدمة ، يهدف هذا البحث إلى المساهمة في تطوير حلول فعالة لتحسين أداء الاستعلام وإدارة قواعد البيانات في عصر البيانات واسعة النطاق .