Graduate College Journal – NU
Vol.20 -2025, No. (03)
ISSN: 1858-6228, http://www.neelain .edu.sd

مجلة الدراسات العليا - جامعة النيلين
المجلد (20) –2025، العدد (03)
الرقم الدولي الموحد للدوريات: 1858-6228

# دراسة طرق تحسين الاستعلام وتأثيرها على نظم إدارة قواعد البيانات (دراسة مقارنة)

**ابتسام الطيب محمد الطيب[1،2*]**
**السماني عبدالمطلب احمد[3]**

[1]قسم علوم الحاسوب - كلية الدراسات العليا - جامعة النيلين - الخرطوم - السودان abtelt@gmail.com
[2]قسم البرمجيات – مركز الحاسوب – جامعة السودان – الخرطوم – السودان
[3]قسم علوم الحاسوب – كلية علوم الحاسوب و تقانة المعلومات – جامعة النيلين – الخرطوم – السودان

## المستخلص

تتناول الدراسة طرق تحسين الاستعلام وكيفية تأثيرها على نظم إدارة قواعد البيانات (DBMS) التي تستخدم للعديد من التطبيقات والخدمات التكنولوجية، يتم التقييم والمقارنة وفقا لمعايير وأسس علمية محددة وتعتبر أمثلية الاستعلام من أهم المعايير التي يتم من خلالها اختيار نظام إدارة قواعد البيانات، DBMS تدعم استعلامات SQL ويمكنها التعامل مع مجموعة ضخمة من البيانات ومعالجة الاستعلامات بكفاءة. ركزت هذه الدراسة على طرق أمثلية الاستعلام التي تتمثل في طريقة التحسين الثابت(Static Optimization)، التي تتكون من إنشاء خطة تنفيذ مثالية. يمكن أن تكون خطط التنفيذ التي تم إنشاؤها بواسطة مُحسّن ثابت دون المستوى الأمثل. الطريقة الثانية طريقة التحسين الديناميكي (Dynamic Optimization)، تستند الأهداف الرئيسية لإدخال "الديناميكية" في تحسين الاستعلام على الاستعداد لاستخدام المعلومات المتعلقة بتوفر الموارد، واستغلال شبه الدقة النسبية لقيم المعلمات، واسترخاء بعض الفرضيات المتشددة وغير الواقعية في سياق ديناميكي مثل (ذاكرة غير محدودة). يتم تطبيق تلك الطرق على DBMS. تتمثل مشكلة الدراسة في عدم وجود دراسات مقارنة شاملة بين طرق أمثلية الاستعلام وفقاً لمعياري الزمن,والتكلفة لمساعدة مديري ومطوري قواعد البيانات في اتخاذهم للقرارات المناسبة و اختيارهم للطريقة المناسبة، ومن ثم تأتى أهمية الدراسة، دراسة مقارنة علمية لطرق الـ(Query Optimization) من خلال إجراء تحليل وصفي لهذه الطرق، أجريت هذه الدراسة لتطوير DBMSباستخدام دراسة مقارنة طرق أمثلية الاستعلام وتأثيرها على DBMS لمساعدة مطوري ومستخدمي قواعد البيانات على اتخاذ القرار في اختيارهم للطريقة المثلى واختيارهم للنظام الذي يفي بمتطلبات التحسين. الهدف من هذه الدراسة هو دراسة مقارنة لطرق تحسين الاستعلام وتوصيف الطرق المثلى لتقديرات التكلفة لتحسين الاستعلام, تحديد أفضل الطرق المستخدمة لتحسين الاستعلام وتقييم ومقارنة هذه الطرق وتأثيرها على نظم إدارة قواعد البيانات Database Management Systems وفقاً لمعياري الزمن,والتكلفة. أجريت التحليلات والمقارنات من خلال عمل جداول لتقييم الأداء والخروج بنتائج تساعد مديري ومطوري ومستخدمي قواعد البيانات في اتخاذهم للقرارات المناسبة واختيارهم للطريقة المثلى للنظام ذات زمن الاستجابة الأقل وزمن التنفيذ الأقل، وتحقيق أمثلية الاستعلام في وجود حلول مثلى لخطط التنفيذ التي تكون دون المستوى الأمثل. من خلال الدراسة والمقارنة وجدنا أن هناك طرق واستراتيجيات وخوارزميات تؤثر علي قواعد البيانات, ووجدنا أن طريقة المحسن الديناميكي أفضل من المحسن الثابت , الديناميكي حقق أمثلية الاستعلام بعدة طرق. وأيضا من خلال النتائج التي توصلنا إليها وجدنا أن طرق و تقنيات تحسين الاستعلام تركز بشكل أساسي على قواعد البيانات المركزية والموزعة. هناك أبعاد مفتوحة للبحث في المستقبل يمكن أن تساعد مصممي ومنفذي و مديري قواعد البيانات، التفكير في استخدام طرق واستراتيجيات وخوارزميات أخرى يمكن تطبيقها على نظم إدارة قواعد البيانات بحيث تلائم النتائج المطلوبة وتفي بالغرض المطلوب وفقاً لمعياري الزمن,والتكلفة.

**الكلمات المفتاحية:** نظم إدارة قواعد البيانات(Database Management Systems(DBMS)، تحسين الاستعلام Query Optimization(QO)، طريقة التحسين الثابت Static Optimization Approach، طريقة التحسين الديناميكي Dynamic Optimization Approach

## مقدمة

يعد تحسين الاستعلام (QO) جانبًا مهمًا في معالجة الاستعلام. إنه نشاط اختيار استراتيجية تنفيذ فعالة لمعالجة الاستعلام ، حيث يوجد العديد من التحويلات المكافئة لنفس الاستعلام رفيع المستوى ، وكذلك وجدنا أن هدف QO يقلل استخدام الموارد ، ويقلل من إجمالي زمن تنفيذ الاستعلام ، تقليل زمن استجابة الاستعلام [1] ، عند وجود مشكلة حسابية مستعصية على

عدد كبير من العلاقات ، يتم تقليل الاستراتيجية المعتمدة لإيجاد أقرب حل أمثل. ومن أهم المتطلبات الوظيفية لنظام إدارة قواعد البيانات هو قدرتها على معالجة الاستعلامات في الوقت المناسب. وينطبق هذا بشكل خاص على التطبيقات المهمة البالغة الأهمية ، مثل التنبؤ بالطقس والأنظمة المصرفية وتطبيقات الطيران ، والتي يمكن أن تحتوي على ملايين وحتى تريليونات من السجلات [2]. هناك نوعان من طرق تحسين الاستعلام( ثابت، وديناميكي). حديثاء استخدمت معظم أنظمة إدارة قواعد البيانات طريقة التحسين الثابت، اقترح باحثون حديثون تحسين دقة قيم الوسائط المستخدمة أثناء اختيار خطة التنفيذ لتصحيح دون المستوى الأمثل تتمثل الحلول في تحسين جودة الإحصاءات على البيانات، تحسين دقة تقدير عوامل الانتقائية(الاختيارية) للمشغل [3], والحل الثاني يركز على الاستعلامات الموزعة واضعين في الاعتبار عامل التكلفة والزمن حيث تعد إدارة قواعد البيانات الموزعة (DDBMS) من بين أهم وأحدث تطورات البرامج في هذا العقد. إنها تتيح وضع الحوسبة والبيانات في بيئة المستخدم بالقرب من أنشطة المستخدم. ترتبط كفاءة أداء DDBMS ارتباطًا عميقًا باستراتيجيات معالجة الاستعلام وتحسينه وتخضع لسيطرة نظام إدارة قاعدة البيانات المركزية (DBMS). في كلا الحلين، يتم تنفيذ خطة التنفيذ المحددة حتى الإنتهاء، مهما كانت التغييرات في بيئة التنفيذ. وبالتالي، فإن محسن الاستعلام كونه وحدة مهمة لنظام ادارة قواعد البيانات له تأثير أكبر على أداء قاعدة البيانات. يقوم بتحليل عدد من الخطط المرشحة التي تم إنشاؤها لاستعلام معين والتي لها مخرجات معادلة ولكن لها تكاليف موارد مختلفة. يختار خطة فعالة من بين خطط هؤلاء المرشحين بأقل تكلفة. على الرغم من أنه تم تنفيذ الكثير من العمل في تحسين الاستعلام، لكن تقدير عدد العناصر في المجموعة لإدارة زمن التحسين وتقدير التكلفة الفعال لا يزال صعباً. أما طريقة التحسين الديناميكي، يمكن تحقيق أمثلية الاستعلام بعدة طرق في بيئات مختلفة مثل أحادي المعالج، الموزع، المتوازي، والواسع النطاق( Uni-processor, Distributed, parallel, Data Integration Systems, and large scale). ووفقًا لكل هذه الطرق فإن لديها القدرة على الكشف عن المثالية الفرعية لخطط التنفيذ وتعديل خطط التنفيذ لتحسين أدائها. والسماح لعملية تحسين الاستعلام أن تكون أكثر قوة فيما يتعلق بأخطاء التقديرات والتغيرات في بيئة التنفيذ. هناك بعض التحليلات والمقارنات النابعة من تقييمات الأداء للاستراتيجيات المقترحة.

***مشكلة الدراسة***

1. المشاكل التي تواجة طريقة المحسن الثابت وخطط التنفيذ التى تم انشاؤها بواسطة محسن ثابت احيانا تكون دون المستوى الأمثل.
2. عدم وجود (تحليل وصفي) بين طرق أمثلية الاستعلام وفقاً لمعايير محددة(الزمن, تقدير التكلفة) لمساعدة مطوري ومستخدمي قواعد البيانات في اختيارهم للطريقة المناسبة.

***أهمية الدراسة***

دراسة مقارنة علمية لطرق تحسين الاستعلام (Query Optimization Approaches) لمساعدة مديري ومطوري قواعد البيانات ومصممي التطبيقات والمستخدمين في اختيارهم للطريقة الأفضل، أجريت هذه الدراسة لتطوير DBMS باستخدام دراسة مقارنة طرق أمثلية الاستعلام وتأثيرها على DBMS والخروج بنتائج تساعد مديرو ومطورو قواعد البيانات ومصممو التطبيقات والمستخدمين على اتخاذهم للقرار المناسب واختيارهم للطريقة الملائمة للعمل مع أنظمة قاعدة البيانات التي تفي بمتطلبات تحسين الاستعلام.

تحسين الاستعلام مهم لأداء قاعدة البيانات العلائقية لأن الاستعلام الأمثل هو التكنولوجيا الرئيسية لكل تطبيق.

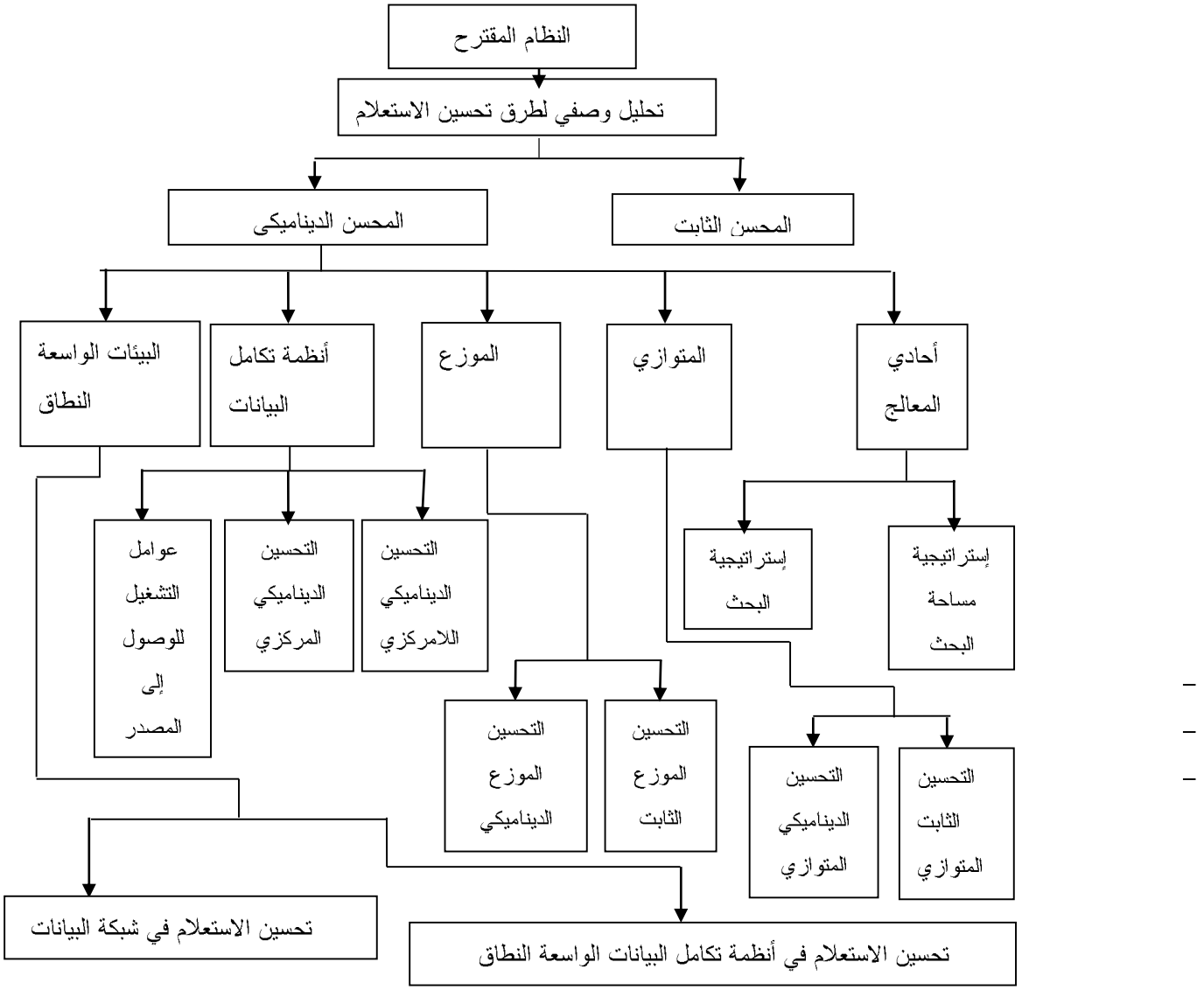
***الهدف من الدراسة***

1- دراسة شاملة لطرق تحسين الاستعلام والتي تتمثل في (التحسين الثابت + التحسين الديناميكي). 2-توصيف الطرق المثلى لتقديرات التكلفة لتحسين الاستعلام. 3-تحديد أفضل الطرق المستخدمة (علمياً وعملياً) .4- تقييم ومقارنة هذه الطرق وتاثيرها على نظم ادارة قواعد البيانات Database Management Systems وفقا لمعايير محددة متمثلة في معيار التكلفة ومعيار الزمن (تقييم زمن تنفيذ الاستعلام وزمن الاستجابة). 5- استخدام عدد من الخوارزميات لتنفيذ عمليات تحسين الاستعلام والمقارنة مثل استراتيجيات مساحة البحث (Search Space strategies) واستراتيجيات البحث(Search Strategies), التي تقوم بالتحليلات والمقارنات وتقييم الأداء.

***الدراسات السابقة***

في الآونة الأخيرة، تناولت العديد من الأبحاث طرق وتقنيات تحسين الاستعلام، وتخزين البيانات والتنقيب عن البيانات، لكنها لم تضع في الاعتبار تنوع مصادر البيانات، وما هي أنواع محسّن الاستعلام الذي يمكن أن يلائم متطلبات كل مصادر البيانات. في الدراسة [4], استخدموا بعض النظريات الأساسية للجبر العلائقي وطرق البحث الاسترشادي الأمثل لتحسين الاستعلام باستخدام خوارزمية (Heuristic Query Optimization Algorithm) ووجدوا أن البحث الاسترشادي الأمثل لتحسين الاستعلام المستخدم في سرعة تنفيذ البيانات كان أسرع بنسبة 23٪ من طلب البحث الأصلي. قدمت الدراسة [5]، مناقشة التقنيات والطرق المختلفة المستخدمة في مجال تحسين الاستعلام وكيفية دمجها أو دمجها في التطبيقات المستقبلية للحصول على أكثر النتائج فاعلية, قدموا

مجموعة تعليمات لتسريع مشغلي قاعدة البيانات الأساسية التي تستخدم في أنظمة قواعد البيانات الحديثة, وأظهروا ميزة (MQO) Multi Query Optimization تقنية لتحسين الأداء الكلي لـ (Hive) Hadoop. في الدراسة [6] ، أظهروا مقارنة بين التكلفة وزمن التنفيذ وزمن الاستجابة بين التحسين القائم على البحث الاسترشادي و تحسين مستعمرة النمل وخوارزميات تحسين التكلفة القائمة على التحسين التكراري. استخلصت في الدراسة [7] ، أن انتشار أنظمة قواعد البيانات العلائقية أنتج مجموعة شاملة من الاستراتيجيات والتقنيات لتحسين أداء الاستعلام، استخدموا cost-based optimizer (CBO) لتقليل تكلفة الاستعلام. في الدراسة [8] ، استخدموا إطار تحسين قائم على التكلفة(CBO) لأجهزة القياس وتنفيذه في GproM، استخدموا طريقة heuristic and cost-based ويمكن تطبيقها على مجموعة واسعة من خطوط أنابيب الأجهزة pipelines.




**الشكل (1) : مخطط تنظيمي للنظام المقترح**

في الدراسة [9]، استخدموا SQL لتكون بمثابة دليل مرجعي لتطوير استعلامات قاعدة البيانات. وضع الباحثون نصائح للحصول على أداء أفضل للاستعلام منها استخدام استراتيجيات تحسين جديدة أكثر تطوراً. في الدراسة [10] ، استخدموا أنواع مختلفة من Swarm Intelligence SI

مثل تلخيص النص القصير (STS)، وقورنت مع التقنيات القائمة على مستعمرة النمل (Ant Colony Optimization ACO).

***– الخوارزميات/ الاستراتيجيات المستخدمة/Algorithms Strategies***

استراتيجيات مساحة البحث: تعتمد على مجموعة التحولات الجبرية التي تحافظ على التكافؤ ومجموعة العوامل الطبيعية المدعومة في المحسن. استراتيجيات البحث: إستراتيجية لحل مشكلة جدولة حل الأنظمة العلائقية أحادية المعالج والتي تميز فئتين من الاستراتيجيات مثل العشوائية، التعدادية. خوارزمية الشجرة الكثيفة، خوارزمية الجدولة التكيفية، خوارزمية البرمجة الديناميكية .

***النظام المقترح***

يتكون النظام المقترح من دراسة مقارنة لطرق تحسين الاستعلام وكيفية تاثيرها على DBMS وفقاً لمعياري الزمن والتكلفة تستند هذه الدراسة على مجموعة من الاستراتيجيات لتنفيذ الاستعلام بجانب خطة جيدة، والعثور على الحد الأدنى من التكلفة لتحسين الاستعلام (التحليل الوصفي لطرق تحسين الاستعلام).

***– المقارنات:***

**الجدول 1: مقارنة استراتيجيات مساحة البحث واستراتيجيات البحث**

| استراتيجيات البحث | | استراتيجيات مساحة البحث | |
|---|---|---|---|
| العشوائية | التعدادية | حجم مساحة البحث | الخصائص/المميزات |
| **تعتمد هذه الاستراتيجيات** على **النهج التحويلي**. تبدأ الاستراتيجيات العشوائية بشكل عام **بخطة تنفيذ أولية** يتم تحسينها بشكل متكرر. يمكن الحصول عليها من خلال **إستراتيجية تعدادية** مثل **الاستدلال** (متعدد المهام) [11]. | **تحدد خوارزمية التعداد للمحسن الخطط** التي يتم تعدادها، تستند بشكل كلاسيكي على **البرمجة الديناميكية**. هناك طريقتان لحل هذه المشكلة تتمثل في الاستراتيجيات **الحتمية** [12]. | **حجم مساحة البحث** هو أهمية شكل الاستعلام )الخطي أو النجم أو الزمرة( **وطبيعة خطط التنفيذ** بسبب حدوثها على حجم مساحة البحث [11]. | يتم **تحديد طبيعة خطط التنفيذ** وفقًا لمعايرين مثل شكل هياكل الشجرة والنظر في الخطط مع المنتجات الديكارتية[11] يمكن عرض **مساحة البحث** كرسم بياني( بعض الوحدات المتنقلة مثل: العقدة، بعض الخطوط بين العقد (الحواف). |
| يمكنه التعامل مع **تحسين الاستعلامات** | لا يمكن **معالجة الاستعلامات** | **الزيادة الهائلة في عدد خطط التنفيذ** حسب | يمكن تقييد مساحة البحث وفقًا لطبيعة **خطط التنفيذ**. إن |
| **المعقدة ( يصبح عدد خطط التنفيذ كبيرًا جدًا)**[11]. | **المعقدة** بسبب **خطط التنفيذ الكبيرة** الخاصة بهم . الاستراتيجيات التعدادية غير كافية في تحسين الاستعلامات المعقدة لأن عدد الخطط كبير[11]. | عدد العلاقات. هذا يوضح صعوبة إدارة مجموعة الحلول التي تكون كبيرة جدًا في بعض الأحيان. لذلك، يجلب هذا ضرورة تكييف إستراتيجية البحث مع خصائص الاستعلام[11]. | الاستعلامات التي تحتوي على عدد كبير من فرضيات الروابط تجعل من الصعب إدارة مساحة البحث المرتبطة التي تصبح كبيرة جدًا [11]. |
| إن تقييم أداء هذه الاستراتيجيات **صعب** للغاية بسبب التأثير القوي، في نفس الوقت، **للمعايير والعوامل العشوائية**. تكمن الصعوبة الرئيسية في اختيار هذه **المعلمات** [11]. | تسمح **بتجاهل الحالات السيئة** من خلال تقديم **الاستدلال** [11]. | شبه **الانضمام**(semi join-) القائم على **الربط**(join) هو تقليل تكاليف الاتصالات **وزيادة حجم مساحة البحث** وتكلفة المعالجة المحلية [13]. | تستهلك **الرسوم البيانية الكبيرة** الكثير من **الذاكرة**، **والأداء** البطيء يتسبب في المزيد من العقد /الحواف [14] . هناك عدد كبير من الفرضيات المرتبطة بـ (join predicates) تجعل **من الصعب إدارة مساحة البحث**، قد اختار بعض الباحثين إزالة أشجار الأدغال(bushy-tree)، ويطلق عليها **مساحة صالحة** تمثل جزءًا كبيرًا من مساحة البحث، وهو الحل الأمثل. تقلل هذه الطرق من فرص الحصول على حلول مثالية[11]. |

**الجدول 2: مقارنة بين طرق تحسين الاستعلام المتوازي (الثابت، الديناميكي)**

| طرق تحسين الاستعلام الثابت المتوازي | | طرق تحسين الاستعلام الديناميكي المتوازي | |
|---|---|---|---|
| التحسين على مرحلة واحدة | التحسين على مرحلتين | التحسين على مرحلة واحدة | التحسين على مرحلتين |
| هناك **طريقتين لربط التجزئة**( hash join) مثل أ/ صلة التجزئة **البسيطة**(simple hash join) ب/ ربط التجزئة **الهجين**( hybrid hash join) [11] | في النهج المكون من مرحلتين، هناك **العديد من استراتيجيات** الجدولة المقترحة لمشغلي خطوط الأنابيب(**pipelined**) [15]. | تشير غالبية العمل إلى أهمية **تحديد درجة التوازي** لعملية الربط وطريقة تخصيص الموارد مثل(المعالجات والذاكرة) [11]. | **طريقة جدولة تكيفية** للفئات النابعة من أفضل **خطة تنفيذ متسلسلة** ممثلة بـ(bushy-tree) أهمية **الخوارزمية جمع الإحصائيات** أثناء تنفيذ الاستعلام. تتوافق الإحصائيات التي تم جمعها مع القيم الحقيقية، حيث يكون التقدير عرضة للخطأ. تُستخدم هذه الإحصائيات لتحسين تخصيص الموارد أو بتغيير خطة التنفيذ [11]. |
| هناك **العديد من الاستراتيجيات** التي تسمح باستغلال قدرات الأشجار العميقة اليمنى( right deep trees) عندما تكون الذاكرة محدودة [11]. | يتم **اقتراح العديد من خوارزميات الجدولة** أي (تخصيص المعالج). وهي مستوحاة من أكبر وقت معالجة (LPT) cessing-Time، تستغل هذه الخوارزميات خطوط الانابيبintra operation partitioned parallelism [11]. | **التعامل مع مشكلة موازنة عبء العمل الديناميكي** للعديد من الاستعلامات المركبة في الـ(hash join). يتم تحديد التوازي أثناء العملية للوصلة (join) وكذلك اختيار معالجات التنفيذ للوصلة (join) بطريقة "متكاملة" من خلال النظر في حالة النظام الحالية. تتميز هذه الحالة | **هذا النهج** ، في الخطوة الأولى ، **يحسن جميع جوانب خطة الاستعلام** بصرف النظر عن اختيار الموقع ؛ الخطة الناتجة من هذه المرحلة هي خطة غير موزعة. يتم جدولة خطة الاستعلام هذه في الخطوة الثانية. وهكذا ينخفض تعقيد التحسين إلى حد كبير [16]. |
| **الاستراتيجيات** التي تسمح باستغلال قدرات الأشجار العميقة اليمنى(right deep trees) عندما تكون الذاكرة محدودة مثل **الاستراتيجية** المسماة (**Static Right Deep Scheduling**) **ومساوئ** هذه الاستراتيجية هي أن عدد الأشجار الفرعية يزداد مع عدد العلاقات الأساسية التي لا يتم تخزينها في الذاكرة. وبالتالي، فإن هذه الطريقة **تقلل** من سلسلة خط الأنابيب **وتزيد** من زمن الاستجابة [11]. | **لتحسين وقت الاستجابة،** يقومون بتطوير نموذج تنفيذي يضمن أفضل مفاضلة (trade-off) بين التنفيذ **المتوازي** (parallel) **ونفقات الاتصال العامة**(communication overhead) [11]. | **تعمل البرمجة الديناميكية** من الأسفل إلى الأعلى من خلال بناء خطط فرعية أكثر تعقيدًا من خطط فرعية أبسط حتى يتم إنشاء الخطة الكاملة. **في المرحلة الأولى،** تبني الخوارزمية خطة الوصول لكل جدول في الاستعلام [12]. **هناك اقتراح توجهات ديناميكية لتخصيص الموارد** في أربع خطوات مطبقة بالترتيب التالي: (أ)الحفاظ على توطين البيانات، (ب)حجم الذاكرة، (ج)تقليل الإدخال/الإخراج، (د)عملية تسلسل شجرة كثيفة(bushy-tree) [11].. | **تعتمد خوارزمية الجدولة التكيفية**( The adaptive scheduling algorithm) على **ثلاثة عناصر**:(أ) **تصنيف المهام** I/O-bound" و "CPU-bound" ، (ب) طريقة **الحوسبة** لنقطة توازن I/O-CPU" " لمهمتين و(ج) **آلية التكيف الديناميكي** لدرجة التوازي للمهمة. يتكون من إيجاد جدولة المهام التي تزيد استخدام الموارد أي (المعالجات والأقراص)، وبالتالي **يقلل من زمن الاستجابة** [11]. |
| **اقتراح طريقتين** أحداهما أ/ **مبنية** على **شجرة مجزأة** ( a segmented tree ) مثل right-deep trees ب/ **مبنية على أشجار متعرجة** [17] ، الهدف من | **تكمن الصعوبة الرئيسية** في تحديد عدد الأجزاء وحجم كل جزء من خلال ضمان أفضل مفاضلة بين | **يبدو أن هذه الطريقة تولد خططًا أفضل** لأنه من الصعب رؤية كيف يمكن بناء الخطة المثلى دون التفكير في موارد وقائية مثل وحدة | **يحدد نوعين من المهام**: المهام المرتبطة بـ **I/O-bound** (محددة بـ المدخلات / المخرجات) والمهام المرتبطة بوحدة المعالجة المركزية **CPU-bound** |

| | | | |
|---|---|---|---|
| الطريقتين هو تجنب التحقيق في مساحة البحث في الأشجار الكثيفة ثم تبسيط عملية التحسين [11]. | عمليات التوازي - النفقات العامة للاتصالات [11]. | المعالجة المركزية والأقراص وشبكة الاتصالات [17]. | (محددة بعدد المعالجات [11] . |

**الجدول3 : مقارنة أمثليه الاستعلام الموزع الثابت والتحسين الديناميكي للاستعلام الموزع**

| تحسين الاستعلام الموزع الثابت | تحسين الاستعلام الديناميكي الموزع |
|---|---|
| **تم تطوير تقنيات تحسين استعلامات الاستعلامات التقليدية** بشكل جيد, ومفيدة بشكل خاص لطلبات البحث التي تعمل على المدى القصير ، على سبيل المثال ، **معالجة المعاملات عبر الإنترنت** ، ولكنها **ليست كافية** للعديد من نطاقات التطبيقات المختلفة اليوم, **حيث تكون الاستعلامات الطويلة المدى شائعة والتي تتراوح من ساعات إلى أيام** [18]. | **تعاني البرمجة الديناميكية من أوقات التنفيذ الطويلة** ومتطلبات الذاكرة الكبيرة جدًا حيث **يزداد حجم العلاقات وعدد الصلات (joins) في الاستعلام** [19]. |
| **تحسين تكاليف الاتصال بين العقد** عن طريق تقليل البيانات المنقولة بينهما [20]. | **البرمجة الديناميكية** تبني الأشجار(trees) فيها بطريقة تصاعدية (bottom-up): يتم اشتقاق الخطط الفرعية الأكثر تعقيدًا من أبسط الخطط، بدءًا من **العقد** الطرفية لخطة الاستعلام النهائية [16]. |
| **تحسين الاستعلام الأمثل** في جميع البيئات [20]. | هناك بشكل **ديناميكي** في الأساس إعادة توجيه الصفوف من خلال المشغلين لتحقيق أوقات استجابة أفضل، تؤثر القرارات على جزء من الخطة المجدولة **ديناميكيًا** [16]. |
| **في التحسين الثابت**، يتم إنشاء الخطط **دون المثالية** بسبب **نقص الموارد**[20] . | **اختيار الموارد** هو مرحلة متكاملة من **تحسين الاستعلام** في معالجة الاستعلام **الموزع**. بغض النظر عن **نهج تعداد الخطة**، مثل البرمجة الديناميكية، والتحسين من خطوتين، **يحتاج مُحسِّن الاستعلام إلى تحديد المورد** الذي سيقيم كل جزء من الاستعلام [16]. |

**الجدول 4 :مقارنة عوامل التشغيل للوصول المقيد(المحدود) إلى المصادر وطرق التحسين الديناميكي (المركزي ، اللامركزي) في أنظمة تكامل البيانات**

| اللامركزي في أنظمة تكامل البيانات | المركزي في أنظمة تكامل البيانات | | عوامل التشغيل للوصول المقيد إلى المصدر |
|---|---|---|---|
| | تعديل خطط التنفيذ على مستوى | | |
| | Inter-operator | Intra-operator | |
| في هذه الطريقة **يصحح المثالية الفرعية لخطط التنفيذ** من خلال لامركزية التحكم [11]. | في هذا المستوى، يتمثل **حل لتصحيح المثالية الفرعية لخطط التنفيذ** في تغيير **جدولة** العملية في وقت التشغيل [11]. | يمكن **تعديل المثالية الفرعية لخطط التنفيذ** خلال هذا المستوى [11]. | **تحتاج خطة التنفيذ** إلى صلة علائقية ( relational join) بين الجداول [11] . |
| **تتطلب خوارزمية ربط التجزئة** Hash join algorithm استقبال جميع الصفوف في المعامل الأول **لبناء جدول التجزئة** قبل بدء خطوة التحقق/الفحص. **يمكن أن يكون وقت إنتاج الصف الأول طويلًا** إذا (أ) كان حجم المعاملات كبيرًا ، أو (ب) عندما يكون معدل وصول البيانات غير منتظم [11]. | هناك **ثلاثة أنواع من التأخير**:التأخير **الأولي**، و **الوصول المتقطع**، و **التسليم البطيء**. يمكن لأسلوب **التحسين الديناميكي** التعامل مع معظم التغييرات في بيئة التنفيذ(**التأخيرات،الأخطاء، والذاكرة غير المتاحة**). تتداخل هذه الطريقة مع مراحل التحسين والتنفيذ **وتستخدم عوامل تشغيل ديناميكية محددة**. في هذه الطريقة، يقوم المُحسِّن بتحويل الاستعلام إلى خطة تنفيذ مشروحة وإنشاء(associated rules) القواعد المرتبطة بنوع(-Event-Condition Action). تحدد هذه القواعد سلوك خطة التنفيذ وفقًا للتغييرات في وقت التشغيل [11]. | **يعتمد هذا المستوى على طريقتين/نهجين** أحدهما يعتمد على **توجيه الصفوف**. يجب أن تكون سياسة توجيه الـ tuple فعالة وذكية من أجل تقليل وقت استجابة الاستعلام. يتم تغليف **جدولة عوامل التشغيل بواسطة التوجيه الديناميكي** لـ tuples. وهناك طريقة/نهج آخر مبني على التقسيم الديناميكي للبيانات [11]. | **يمكن تصميم الجداول** بمفهوم "أنماط الربط ( binding patterns)" وأنماط الربط يمكن إرفاقها بالجدول العلائقي لوصف قيود الوصول إليها بسبب أسباب السرية أو قضايا الأداء [11]. |
| في هذه الطريقة **لحل مشكلة المعاملات** ذات الأحجام الكبيرة أو مع **الاستعلامات المكونة من عدة** | في هذه الطريقة **التحقق من شروط معينة** (مثل **مقارنة** أحجام العلاقات المؤقتة الحالية مع تلك المقدرة خلال التجميع) عندما تحدث الأحداث (على سبيل المثال، **التأخير**، | **يعتمد تقسيم البيانات الديناميكية** على مزيد من المعلومات الإجمالية لإنشاء الخطط الجديدة. تمثل **المجموعات** التي | على الرغم من أن النموذج يمكن أن يكون **المصادر المقيدة** مع إضفاء الطابع |

| | | | |
|---|---|---|---|
| **وصلات**، يتم نقل أجزاء من **جداول التجزئة** الموجودة في الذاكرة نحو مساحة تخزين ثانوية. **عندما تصبح الذاكرة مشبعة**، يتم اختيار قسم من أحد الجدولين ليتم نقله نحو مساحة التخزين الثانوية [11]. | **الذاكرة** غير متوفرة) تبدأ الإجراءات (مثل إعادة **تخصيص الذاكرة**، إعادة **الجدولة** أو إعادة **التحسين**) [11]. | تتم معالجتها بواسطة كل خطة مستخدمة(بيانات مقسمة). في هذه الآلية عندما يتم استبدال خطة التنفيذ ، يتم إنتاج تقسيم جديد للبيانات. تنتج كل خطة تنفيذ مستخدمة جزءًا من النتيجةالإجمالية [11] . | الرسمي على "أنماط الربط" ، نظرًا لقيود الوصول إلى **المصادر**، لا يمكن استخدام **عوامل تشغيل معالجة الاستعلام**، مثل المسح ألعلائقي والانضمام ألعلائقي ( relational scan and relational join). **هناك حاجة لمشغل مسح جديد قادر على التعامل مع المصادر المقيدة** [11]. |

**الجدول 5 : مقارنة أمثلية الاستعلام في أنظمة (تكامل البيانات الواسعة النطاق، و شبكة البيانات)**

| تحسين الاستعلام في أنظمة تكامل البيانات الواسعة النطاق | تحسين الاستعلام في أنظمة شبكة البيانات |
|---|---|
| **تعريف**: تكامل البيانات الناشئة/الناتجة عن عدة مصادر.<br>**الأهداف**: الوصول الفعال والشفاف/الواضح والموحد لمصادر البيانات [13]. | **التعريف**: تنسيق البنية التحتية للحوسبة الموزعة. تقاسم الموارد وحل المشكلات المنسقة بطريقة ديناميكية. **الأهداف**: طرق تخصيص و جدولة الموارد الفعالة، تكامل بُعد التوازي والنظر في عدم التجانس، وتوفير بيئة قوية[13]. |
| **الخصائص**:(التوزيع، عدم التجانس، الحكم الذاتي(الاستقلالية) [13]. تعمل أنظمة تكامل البيانات على توسيع طريقة قاعدة البيانات الموزعة ليشمل مصادر بيانات متعددة ومستقلة وغير متجانسة من خلال توفير موحد الوصول(نفس واجهة الاستعلام للقراءة فقط لجميع المصادر) [11]. | هناك **خاصيتان**: مقياس كبير(بما في ذلك (الاستقلالية، عدم التجانس)) وعدم استقرار الأنظمة ("ديناميكيات العقد")<br>**(استغلال الموارد المتاحة**: وحدة المعالجة المركزية والذاكرة والشبكة وعرض النطاق الترددي للإدخال / الإخراج ومصادر البيانات والخدمات) [13]. |
| **الافتراض**: استقرار الأنظمة (من Uni-proc DBMS إلى تكامل البيانات على نطاق واسع) افتراض قوي في بيئة واسعة النطاق [13]. | **التطور**:العديد من مشكلات إدارة البيانات في أنظمة الشبكة(اكتشاف الموارد وتحديدها، ومعالجة الاستعلام والتحسين، وخدمات المراقبة، التكرار التخزين المؤقت، نماذج التكلفة، البيانات الذاتية وقضايا الأمان)تركز في هذه الطريقة على معالجة الاستعلام والتحسين[13]. |
| **القيد**:عدم استقرار الأنظمة(يمكن للعقدة الانضمام أو المغادرة أو الفشل في أي وقت(ديناميكية العقد)[13] . | **القيد**: أشكال التوازي (التوازي المقسم، المستقل، الموازي للأنابيب)، تكلفة الاتصالات وموازنة الحمل[13]. |

***النتائج:***

من خلال نتائج الدراسة والتحليل الوصفي لطرق تحسين الاستعلام وجد حديثاً أن معظم أنظمة إدارة قواعد البيانات (DBMS) استخدمت طريقة التحسين الثابت. تفترض جميع الطرق ، باستخدام هذه الطريقة ، أن قيم المعلمات / المستخدمة لتوليد خطة التنفيذ تكون دائمًا صالحة أثناء تنفيذها. في الواقع ، يمكن أن تصبح قيم هذه المعلمات غير صالحة أثناء التنفيذ لأسباب عديدة (أخطاء تقدير ، عدم توفر الموارد). وفقًا للأسباب السابقة، يمكن أن تكون خطط التنفيذ التي تم إنشاؤها بواسطة مُحسّن ثابت دون المستوى الأمثل. على الرغم من أنه تم تنفيذ الكثير من العمل في تحسين الاستعلام ، لكن تقدير عدد العناصر في المجموعة لإدارة زمن التحسين وتقدير التكلفة الفعال لا يزال صعباً . وأيضا وجدنا أن طريقة التحسين الديناميكي، عدلت من خطط التنفيذ دون المستوى الأمثل. الديناميكي حقق أمثلية الاستعلام بعدة طرق في بيئات مختلفة مثل أحادي المعالج : **الجدول (1)** يمثل نتائج مقارنة بين استراتيجيات مساحة البحث, واستراتيجيات البحث, المتوازي : **الجدول (2)** يمثل نتائج مقارنة بين طرق تحسين الاستعلام المتوازي (الثابت، الديناميكي), الموزع : **الجدول (3)** يمثل نتائج مقارنة بين أمثليه الاستعلام الموزع (الثابت، الديناميكي), تكامل البيانات : **الجدول (4)** يمثل نتائج مقارنة بين عوامل التشغيل للوصول المقيد (المحدود) إلى المصادر وطرق التحسين الديناميكي (المركزي، اللامركزي) في أنظمة تكامل البيانات, البيئات الواسعة النطاق : **الجدول (5)** يمثل مقارنة بين أمثلية الاستعلام في أنظمة (تكامل البيانات الواسعة النطاق، و شبكة البيانات). من خلال **الجدول (3)** و **الجدول (4)** توصلنا إلى أن طريقة قواعد البيانات الموزعة تعالج حوالي عشرات قواعد البيانات الموزعة بينما يمكن لطريقة نظام تكامل البيانات أن يصل إلى مئات مصادر البيانات. هذه الطرق كشفت عن المثالية الفرعية لخطط التنفيذ وعدلت من خطط التنفيذ لتحسين الأداء. وسمحت لعملية تحسين الاستعلام أن تكون أكثر قوة فيما يتعلق بأخطاء التقديرات والتغيرات في بيئة التنفيذ. من خلال النتائج أعلاه وجد أن طريقة المحسن الديناميكي أفضل من طريقة المحسن الثابت.

***المناقشة:***

- من خلال تحليلات المقارنة لزمن وتقدير تكلفة التنفيذ وجدنا أن الإستراتيجية التعدادية تتبع البرمجة الديناميكية، لكنها لا تستطيع معالجة الاستعلامات المعقدة بسبب خطط التنفيذ الكبيرة. الإستراتيجية العشوائية يمكنها التعامل مع تحسين الاستعلامات المعقدة ( يصبح عدد خطط التنفيذ كبيرًا جدًا) يتم تحسينها بشكل متكرر بتطبيق مجموعة من قواعد التحول (الجدول (1)).
- تعتمد مساحة البحث على مجموعة التحولات الجبرية, يستخدم المُحسِّن العديد من العروض التوضيحية للاستعلام. تقسم خطة مساحة البحث إلى مساحات فرعية, يتم إنتاج الخطة استنادًا إلى وسائط النظام الحالية وأداء الخطط التي تم إنشاؤها سابقًا للاستعلام نفسه بحيث يمكن إجراء تحسين إضافي (الجدول (1)).
- في المتوازي الثابت, لتحسين زمن الاستجابة ، يقومون بتطوير نموذج تنفيذي يضمن أفضل مفاضلة (trade-off) بين التنفيذ المتوازي (parallel) ونفقات الاتصال العامة. المتوازي الديناميكي يعتمد على خوارزمية الجدولة التكيفية (الجدول (2)).
- لتعظيم الاستفادة من الموارد يقترح طريقة التكيف الديناميكي لدرجة التوازي للمهمة وفقا لطرق التوزيع المنفذة مثل (جولة روبن (round-robin) ، الفاصل الزمني(interval)). يتم استخدام هذه الطريقة في الجدولة التكيفية بحيث يعمل النظام دائمًا على نقطة توازن IO-CPU [11] (الجدول (2)).
- لا تتطور الاستراتيجيات الديناميكية لأن مركزية القرار (التي يتخذها المُحسن) هي الاختناق (bottleneck) وعدم الدقة والتقادم في التقديرات (الإحصائيات واختيار الفرضيات وتكلفة المشغلين) وعدم توفر الموارد (وحدة المعالجة المركزية والذاكرة والشبكة ومصادر البيانات) (الجدول (2)).
- اقترح الباحثون جدولة الخوارزميات التي تستغل/ تستخدم فقط pipeline parallelism(POT Pipelined Operator Tree Scheduling) ، ثم يوضحون كيفية تمديد هذه الخوارزميات لتأخذ في الاعتبار التوازي أثناء العملية وتكاليف الاتصال [11] (الجدول (2)).
- في الموزع الديناميكي تعاني البرمجة الديناميكية من أوقات التنفيذ الطويلة ومتطلبات الذاكرة الكبيرة جدًا بينما في الموزع الثابت تم تطوير تقنيات تحسين استعلامات الاستعلامات التقليدية بشكل جيد, ومفيدة بشكل خاص لطلبات البحث التي تعمل على المدى القصير( الجدول (3)).
- تحسين الاستعلام مهمة صعبة في البيئة الموزعة بسبب العديد من العوامل مثل توزيع البيانات ، وسرعة قناة الاتصال ، والفهرسة ، وتوافر الذاكرة ، وحجم قاعدة البيانات ، وتخزين النتيجة الوسيطة ، وخط الأنابيب(pipeline) وحجم نقل البيانات حيث يستخدم العديد من طرق التحسين الموزعة البرمجة الديناميكية مع بعض خوارزميات التعداد الأخرى لإجراء اختيار الخطة ( الجدول (3)).
- تعمل طرق التحسين على تصحيح المثالية الفرعية لخطط التنفيذ من خلال لامركزية التحكم بينما يتمثل حل لتصحيح المثالية الفرعية لخطط التنفيذ في تغيير جدولة العملية في وقت التشغيل من خلال المركزية وفي عوامل التشغيل للوصول المقيد إلى المصدر تحتاج خطة التنفيذ إلى صلة علائقية بين الجداول ( الجدول (4)).
طبيعة صنع القرار (بواسطة المُحسِّن): التحكم المركزي (العملية الفريدة (المُحسِّن) هو المسئول عن الإشراف والكشف وتعديل خطة التنفيذ والتحكم اللامركزي (الكشف والتعديل من خلال عدة عمليات) ( الجدول (4)).
- الهدف من الواسعة النطاق في أنظمة تكامل البيانات الوصول الفعال والموحد لمصادر البيانات والهدف من شبكة البيانات جدولة الموارد الفعالة، تكامل بُعد التوازي والنظر في عدم التجانس، وتوفير بيئة قوية ( الجدول (5)).
- الاستقلالية، عدم التجانس وعدم استقرار الأنظمة ("ديناميكيات العقد") في الواسعة النطاق وشبكة البيانات( الجدول (5)).

- في بيئة موزعة على نطاق واسع، ينخفض أداء تحسين الاستعلام في نظام تكامل البيانات(عدد الرسائل المهمة نسبيًا على شبكة ذات نطاق ترددي منخفض وتأخير في زمن الانتقال, الاختناق) (الجدول (5)).
- تبني إستراتيجية الجشع (greedy) خطة واحدة فقط من هذا النوع باستخدام البحث العميق (depth-first) ، بينما تبني البرمجة الديناميكية (dynamic Programming) جميع الخطط الممكنة على نطاق واسع (breadth-first) [8].
- حديثاً، أدى التقارب بين تقنيات الشبكات وخدمات الويب إلى قيام الباحثين بتطوير واجهات شبكة موحدة. تعتبر بنية خدمات الشبكة المفتوحة واحدة من أكثر المعايير المعروفة المستخدمة في الشبكات. يتم تطوير العديد من التطبيقات باستخدام معايير هندسة خدمات الشبكة المفتوحة Open Grid Services Architecture.

***الخلاصة***

قدمت الدراسة تحليل وصفي مفيد ودراسة مقارنة لطرق امثلية الاستعلام لمساعدة مبرمجي ومطوري ومديري قواعد البيانات في اختيارهم للطريقة المناسبة، من خلال الدراسة والتحليل وجدنا أن طريقة المحسن

الديناميكي افضل من المحسن الثابت من حيث الزمن والتكلفه، الديناميكي حقق امثلية الاستعلام. بعدة طرق أحادي المعالج :استقلالية البيانات. الموازي: الأداء العالي وتوافر البيانات. الموزع: الموقع والتجزؤ وشفافية التكرار. أنظمة تكامل البيانات: وصول موحد لمصادر البيانات، عدم التجانس، الاستقلالية. نظام شبكة البيانات: استغلال الموارد المتاحة. واسعة النطاق : تعتمد جودة التنفيذ وتكلفة التحسين على جودة الاختيار, تركز طرق و تقنيات تحسين الاستعلام بشكل أساسي على قواعد البيانات المركزية والموزعة.

**التوصيات**

هناك أبعاد مفتوحة للبحث المستقبلي كدليل يمكن أن يساعد مصممي قواعد البيانات أو منفذي نظام قاعدة البيانات,

- ✓ التفكير في طرق واستراتيجيات وخوارزميات أخرى يمكن تطبيقها على نظم إدارة قواعد البيانات بحيث تلائم النتائج المطلوبة وتفي بالغرض المطلوب.
- ✓ التفكير في طرق واستراتيجيات وخوارزميات أخرى لتحسين الاستعلام في البيئة الموزعة لان تحسين الاستعلام مهمة صعبة في البيئة الموزعة بسبب العديد من العوامل مثل توزيع البيانات ، وسرعة قناة الاتصال ، والفهرسة ، وتوافر الذاكرة ، وحجم قاعدة البيانات ، وتخزين النتيجة الوسيطة ، وخط الأنابيب(pipeline) وحجم نقل البيانات حيث يستخدم العديد من طرق التحسين الموزعة البرمجة الديناميكية مع بعض خوارزميات التعداد الأخرى لإجراء اختيار الخطة.

1. ***المصادر والمراجع***


[1] S. Sumathi & S. Esakkirajan, "Fundamentals of Relational Database Management Systems", 2007- on Amazon.com.

[2] M.L. Rupley & Jr.Indiana University at South Bend,"Introduction to Query Processing and Optimization". http://www.cs.iusb.edu/technical_reports/TR-20080105-1. Pdf.

[3] M.Khan & M. N. A. Khan, "Exploring Query Optimization Techniques in Relational Databases " , 3, June, 2013.

[4] I N E Indrayana1, N M Wirasyanti D P2&I KG Sudiartha3 " Heuristic query optimization for query multiple table and multiple clausa on mobile finance application" The 2nd International Joint Conference on Science and Technology (IJCST) (2017) IOP Publishing.

[5] Rini John, Department of Computer Engineering of PIIT Mumbai University, New Panvel, India Nikita Palaskar University of Memphis," A Survey of Various Query Optimization Techniques " International Journal of Computer Applications (0975–8887) September 2017

[6] Ajay Wagh & Varsha Nemade "Query Optimization using Multiple Techniques" International Journal of Computer Applications (0975 – 8887) 3, April 2017.

[7] Derek Colly & Dr. Clare Stanier , "Identifying New Directions in Database Performance Tuning ",2017.

[8] Xing Niu & Raghav Kapoor & Boris Glavic & Dieter Gawlick & Zhen Hua Liu & Venkatesh Radhakrishnan , "Provenance-aware Query Optimization", 2017.

[9] Luqman Hakim & Suryanto Nugroho & Sigit Hadi Waluyo, "Experiment of Query Optimization Techniques to get the Efficient Query", 2017.

[10] Mohamed Atef Mosa & Arshad Syed Anwar and Alaa Hamouda, "A survey of multiple types of text summarization based on swarm intelligence optimization techniques", 2018.

[11] F.Morvan & A.Hameurlain, "Evolution of Query Optimization Methods", 2009.

[12] Ankush Sawarkar, " Query Processing and Query Optimization in Distributed Database: A Survey", 2014.

[13] Prof.A.meurlain , "Evolution of Data Management Systems: from Uni-processor to Large-Scale Distributed Systems",2017.

[14] Dan. Bader & Paul. Tozour, "Pathfinding: Search Space Representations", Chapter 2.1.ppt.

[15] W.Hasan & R.Motwani, "Optimization Algorithms For Exploiting The Parallelism–Communication Tradeoff in Pipelined Parallelism", pdf.

[16] F.Yuanyuan & Mi Xifeng, "Distributed Database System Query Optimization Algorithm Research", IEEE, 2011.

[17] A.Hameurlain & F.Morvan, "Evolution of Query Optimization Methods", 2009.

[18] Kenneth W. Ng & Z.Wang and R.R.Muntz & S.Nittel, "Dynamic Query Re-Optimization", pdf .

[19] Tansel Dökeroğlu, Ahmet Coşar , "Dynamic Programming with Ant Colony Optimization Metaheuristic for Optimization of Distributed Database Queries",2014.

[20] M.Khan & M. N. A. Khan ," Exploring Query Optimization Techniques in Relational Databases ", 2013.

.