إستراتيجية لتعديل البيانات في نظام قواعد بيانات موزع
Strategy of Update The Data in Distributed Database System

د/طاهر علي الرشاحي

أستاذ الحاسوب المساعد – كلية هندسة وعلوم الحاسوب – جامعة الحديدة

الملخص :

في هذا البحث سوف يتم تناول مشكلة تعديل البيانات في نظام قواعد بيانات موزعة يحتوي على ملفات لها عدة نسخ داخل هذا النظام. توجد عدة استراتيجيات سبق أن قدمت حلول جزئية حول هذه المشكله و كانت تقتصر على تعديل بيانات مركزية بدون نسخ ،وفي معظم الأحيان تعتمد على إستراتيجية التعديل الآجل أو التعديل في الزمن الحقيقي . في إطار نموذج لنظام قواعد بيانات موزعه يحتوي على بيانات مكررة (عدة نسخ) في أكثر من نقطة داخل شبكة كمبيوتر محلية، سوف يتم استعراض إستراتيجية لتعديل البيانات في هذه النسخ، بما يضمن توافق بياناتها.

الإستراتيجية المقدمة شملت مزايا كل من التعديل الآجل وكذا التعديل الحاصل في الزمن الحقيقي وتضمن إمكانية تعديل البيانات في النقاط الخارجة عن العمل داخل النظام الموزع بمجرد أن تدخل فيه من جديد ، وبذلك نحصل بشكل مستمر على بيانات متوافقة داخل النظام ترفع من درجة مستوى الوثوقيه فيه.

الكلمات الداله : استراتيجية تعديل البيانات في الأنظمة الموزعة، نظم قواعد البيانات الموزعة، تحديد مواقع الملفات وعدد النسخ لكل منها.

Key Words: Strategies of Update in distributed systems, Distributed database systems (ddbs), Replication data in (ddbs), data catalog.

١- مقدمة :

أنواع العمليات في الأنظمة الموزعه تتمثل في عملية القراءة والكتابة. عملية القراءة يمكن أن تكون قراءة محلية أو قراءة "عن بعد" بحسب طبيعة الاستعلام وأماكن توزيع البيانات داخل النظام الموزع، بينما عملية الكتابة يمكن أن تكون إحدى هذه العمليات:

١- إضافة بيانات

٢- تعديل بيانات

٣- حذف بيانات

هذه العمليات يمكن تنفيذها محليا على بيانات الجهاز الذي قام بالعملية أو "عن بعد" على بيانات جهاز أخر. لاشك أن هذه العمليات تؤثر بشكل أو بأخر على كفاءة النظام ووثوقيته وذلك من خلال :

- ارتفاع تكاليف الاتصالات نتيجة القراءة "عن بعد" أو الكتابة أيضا.
- إمكانية سقوط بعض نقاط النظام والتي تكون بياناتها هامه لمستخدمين آخرين.
- قدرته على الاستمرار في تنفيذ وظائفه.

هذه العوامل وغيرها دفعت مصممي نظم قواعد البيانات الموزعة إلى اختيار أخر يزيد من وثوقية النظام ويعمل على إمكانية استمراره في تنفيذ وظائفه دون توقف ، الاختيار يتمثل في تطبيق آلية تعدد النسخ للملفات التي يتم الاستعلام عنها "عن بعد " بشكل دائم [**1**]. هذه الآلية ساهمت في:

- انخفاض تكاليف الاتصالات داخل الشبكة.
- زيادة سرعة الشبكة في تنفيذ المهام نتيجة الحد من الاستعلامات "عن بعد".
- سهوله الوصول إلى البيانات.
- استمرارية العمل بغض النظر عن وجود نقاط في النظام قد لا تعمل.
- ارتفاع مستوى الوثوقيه

مع كل المميزات السابقة الذكر إلا أن آلية تعدد النسخ تحمل إلينا مشاكل جديدة بحاجة إلى وجود إجراءات تعمل على حلها ، لقد لوحظ من خلال التجارب العملية **[2,3]** انه كلما زادت عدد النسخ قلة التكلفة فيما يتعلق بعملية القراءة كونها تنفذ محليا ،ولكن ما أن تبداء عملية التعديل في بيانات إحدى النسخ حتى ترتفع هذه التكاليف إلى أعلى مستوى نتيجة ضرورة نشر هذه التعديلات إلى جميع النسخ حتى يكون هناك توافق في البيانات.

مشكلة تحديد عدد النسخ لكل ملف وكذلك تحديد مواقعها تم بحثه في [٢٣]، حيث قدم نموذج ديناميكي يقوم بزيادة عدد النسخ عندما تكون الاستعلامات داخل النظام خاصة بعملية القراءة لكي نحصل على اقل التكاليف، وعندما تكون العملية "تعديل " فان التعديلات تحدث فقط على عدد معين من النسخ تم تحديده مسبقا بخوارزمية معينه تكفل الحصول على اقل التكاليف داخل النظام، بشرط أن يكون اقل عدد للنسخ هو

نسختين. وان كان هناك ارتباط بين عدد النسخ وعملية التعديل ،فان تأثير هذه العملية ليس فقط في رفع تكاليف الاتصالات وإنما تتعدى في البحث و الحصول على استراتيجية تضمن أن التعديل الحاصل في إحدى النسخ يجب أن ينتشر ليصل إلى جميع النسخ بما يحقق :

- الالتزام بخصائص الحركة "العملية"[٧] من حيث نقل البيانات من وضع صحيح إلى وضع صحيح آخر وكذا استمرارية اثر التعديل الحاصل أي أن يكون دائم .
- أن تكون البيانات متوافقة داخل جميع النسخ.

إن عملية التعديل تعتبر عملية غالية الثمن ويجب اختيار الاستراتيجية التي تساعد في خفض التكاليف وزيادة وثوقية النظام .

لكي نتمكن من عرض الاستراتيجية المقدمة، فإننا سوف نتطرق أولا إلى معرفة طرق آلية التعديل حيث انه قد أمكن تصنيف أنواع التعديل في أنظمة قواعد البيانات ألموزعه إلى نوعين :

- التعديل الآجل والذي يتم في أي وقت بعد انتهاء العملية بحسب ما يحدده مصمم نظام قواعد البيانات الموزع " في أخر الدوام- كل يومين .... الخ".
- التعديل في الزمن الحقيقي ويكون وقت حدوث العملية.

٢- الخلفية التاريخية للمشكلة

يمكن التطرق إلى مجموعة من الطرق المستخدمة في تعديل البيانات ذات النسخ في أنظمة قواعد البيانات ألموزعه،أشيرا إليها في كل من [**4,5**] منها ما يستخدم حاليا في بعض نظم قواعد البيانات " التجارية". من هذه الطرق طريقة **Sneaker net** والتي تعتمد على تعديل البيانات في إحدى النسخ وبعد الانتهاء يتم نقل البيانات المعدلة على قرص ممغنط أو أي وسيلة تخزين خلال فترة زمنية معينه خلال ٢٤ ساعة يعدل فيها باقي النسخ. وتعتبر هذه الطريقة غير مجدية في حالة تأخر وصول البيانات المعدلة حيث انه قد يقدم احد المستخدمين على استخدام البيانات الغير معدله مما قد يؤدي إلى عدم توافق البيانات ويترتب عليه اتخاذ قرارات غير صحيحة. أو قد يطول فترة الانتظار إلى حين وصول البيانات المعدلة.

في نظام قواعد البيانات اوراكل (**Oracle**) قدمت عدة استراتيجيات لتعديل البيانات منها استخدام الصور اللحظية (**Snapshots**) والتي فقط تستخدم للقراءة ، عند حدوث أي تعديل في الملف الأساسي يقوم النظام بعملية تحديث للصورة اللحظية لدى المحطة الطرفية ومن هنا نجد أن التعديل قد وقع فقط من طرف واحد. وفي حالة تم تعديل الصورة اللحظية فإن التعديل يرسل فقط إلى الملف الأساسي وفي هذه اللحظة تكون

عدد النسخ التي يجب تعديلها واحدة مما قد يجعل الأمر اقل تعقيدا ،ومع ذلك فان المشكلة تكمن في حالة كان الجهاز الذي يحوي على الملف الأساسي خارج عن العمل عندئذ يصبح من المستحيل عمل أي تعديل على الصور اللحظية.

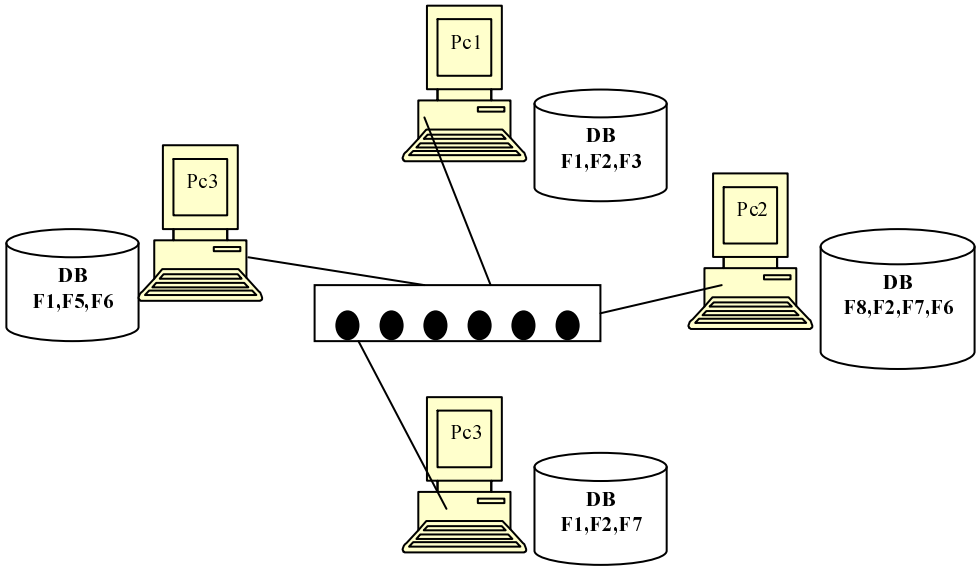
أيضا هناك طريقه أخري تقوم على تعديل النسخ التي تمثل الملف الرئيس وذلك بعد أن يقوم مدير النظام بتحديدها فانه يتم إرسال التعديلات إليها عند حدوث العملية ومن عيوب هذه الطريقة انه إذا وجد أن احد الأجهزة التي توجد لديه نسخه ،غير قابل للتعديل لسبب ما فإن التعديل لن يتم مما قد يؤدي إلى إلغاء العملية.

٣- إستراتيجية لتعديل البيانات في نظام قواعد البيانات الموزع

من خلال السيرة التاريخية لهذه المشكلة، نجد إننا بحاجه إلى إستراتجية تعمل على تجاوز هذه المشكلات والإستراتيجية المقترحة في هذا البحث سوف تقدم حلول تضمن أن تكون البيانات في النظام الموزع في حالة توافق دائم، تمنع عدم تناقضها وتساهم في رفع مستوى الوثوقية بالنظام الموزع.

يتم ذلك من خلال الجمع بين كل من آلية التعديل في الزمن الحقيقي وكذا آلية التعديل في الزمن الآجل عن طريق إضافة مفهوم جديد إليهما يتمثل في أنشاء ملف يعتبر بمثابة "صندوق بريد " لكل نسخة يعمل هذا الصندوق على استقبال البيانات المعدلة في حالة لم تنجح آلية التعديل في الزمن الحقيقي ،كون هذه النقطة تعتبر خارج النظام لسبب من الأسباب، وحينها فان دور صندوق البريد يأتي عند دخول هذه النقطة ضمن النظام حيث يعمل على تفعيل آلية التعديل الآجل من خلال إرسال محتوياته من البيانات إلى النسخ التي لم تتمكن من تعديل بياناتها.

شكل رقم (١) نظام قواعد بيانات موزع



يتم إنشاء ملفات صندوق البريد بعدد النسخ كي تعمل على استقبال التعديلات وفي حالة فشل التعديل في الزمن الحقيقي في أي نقطة من نقاط النظام التي تحوي على النسخ يتم تفعيل التعديل في الزمن الآجل من خلال تحديث بيانات تلك النسخ عن طريق نقل محتويات صندوق البريد.

لنقل محتويات صندوق البريد من السجلات المعدلة يمكن أن تستخدم تقنيات **XML Distributed Systems Design** على هيئة رسائل من نوع **Point-to-point** [6] ، يتم تبادلها داخل نظام قواعد البيانات الموزعة وإضافتها إلى النسخ التي لم تعدل.

٣-١- شروط لنجاح عملية التعديل :

١ – عند عملية التعديل في الزمن الحقيقي ترسل التعديلات إلى جميع النسخ ،وفي حالة وجود نسخه أو أكثر خارج النظام فان عملية التعديل تنفذ طالما أن هناك نسختين كحد أدنى .

٢ – ترسل التعديلات إلى صندوق البريد الخاص بكل نسخه.

٣– الأجهزة التي نجحت في عملية التعديل لا تستقبل صناديق بريدها أي بيانات معدله .

٤ – تحصل النسخ الغير معدلة على التعديلات الجديدة بمجرد دخولها في النظام بأخذها من صندوقها البريدي باستخدام آلية التعديل الآجل.

٥ – تحذف محتويات الصندوق بعد انتهاء العملية.

٤- آلية تعديل البيانات في نظام قواعد البيانات الموزعة

عند الشروع في عملية التعديل على البيانات في إحدى الملفات التي لها عدد من النسخ داخل النظام تنفذ الخطوات التالية كما في الشكل رقم (٢):-

١ – إرسال رسالة إلى قاموس البيانات في النظام الموزع للتأكد من وضع الملف هل هو في حالة استخدام أم لا.

٢ – إذا كان في حالة استخدام ينتظر ثم يعاود إلى النقطة (١).

٣- توضع علامة على حالة الملف وكذلك على جميع نسخة انه في حالة استخدام .

٤ – تتم عملية التعديل على البيانات داخل الملف .

٥ – في حالة أن جميع الأجهزة التي تحوي على النسخ فاعلة في النظام تنفذ عملية التعديل حتى النقطة (٨).

٦ – يتم تنفيذ التعديل إذا وجد على الأقل نسختين في النظام وقت الحدث كحد ادني.

٧- إرسال البيانات المعدلة إلى صندوق بريد كل نسخة لم تعدل .

٨- عند الانتهاء ترسل رسالة إلى قاموس البيانات لإزالة علامة الاستخدام (النقطة ١١)

٩ – الأجهزة التي لم تعدل بياناتها عندما تدخل في النظام تعمل على تحسس صناديق بريدها وتقوم بإضافة محتوياته ويسمى التعديل الآجل.

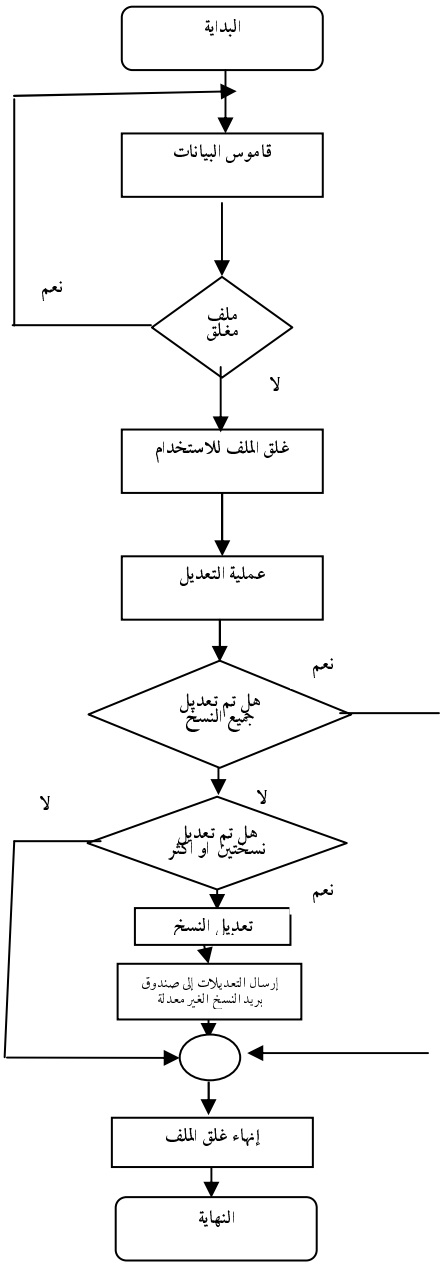
١٠ – عند الانتهاء يتم حذف محتويات هذه الصناديق.

١١ – انتهاء التعديل

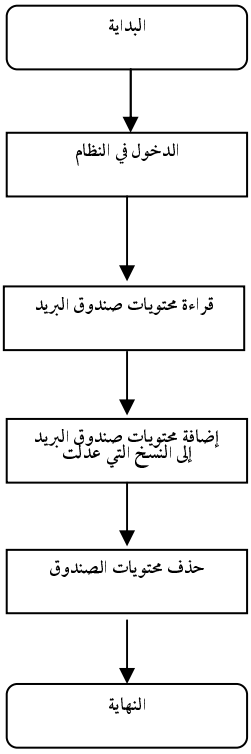
بهذه الإستراتيجية استطعنا أن نجمع بين مزايا التعديل في الزمن الحقيقي وكذلك الاستفادة من التعديل الآجل من خلال استخدام تقنية صناديق البريد التي تتمثل في ملفات تجمع فيها البيانات التي تم تعديلها ولم تستقبل بعد من قبل تلك المحطات التي توجد فيها النسخ، ونستخدم التعديل الآجل لنشرها بما يحافظ على توافق البيانات باستمرار وفي نفس الوقت لقد ساعدت هذه الإستراتيجية على استمرار وظائف النظام كون أن هناك أكثر من مصدر للحصول على البيانات المعدلة.

الشكل رقم (٣) يمثل تنفيذ جزء من خوارزمية تعديل البيانات في الزمن الحقيقي بواسطة أوامر لغة **SQL** في وسط نظام إدارة قواعد البيانات **Oracle**

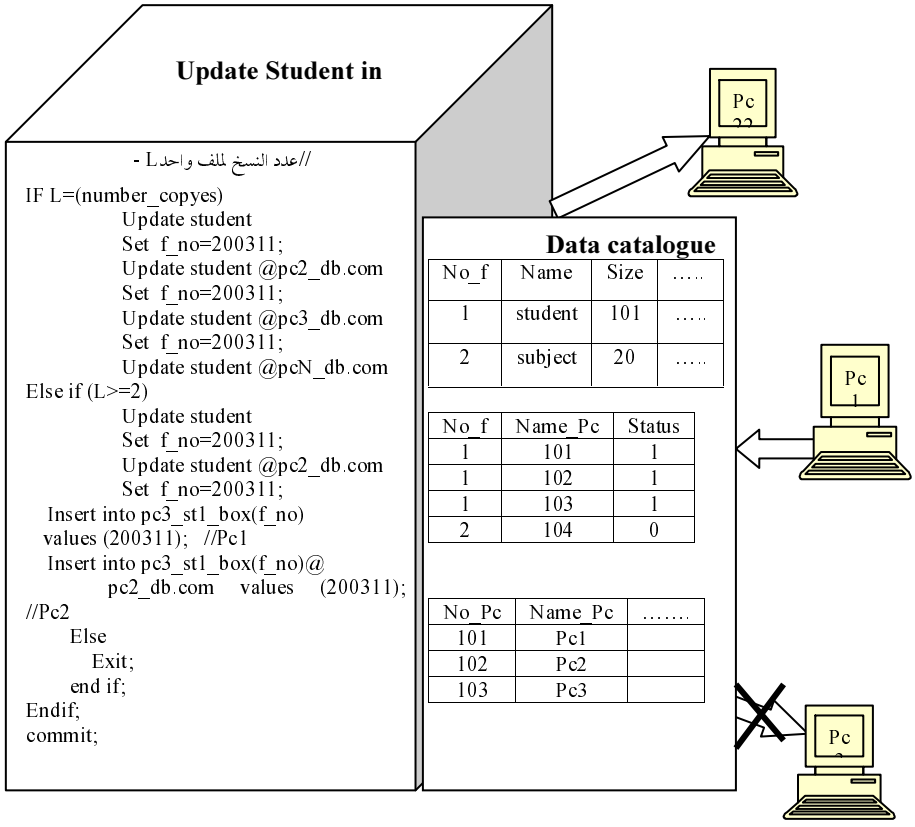
ا - وقت حدوث العملية



ب - تعديل اجل



الشكل 2 يوضح خوارزمية التعديل



الشكل رقم (٣)

٥- التوصيات

إن احد جوانب القوة في أنظمة قواعد البيانات الموزعة هو استمرارية عملها مما جعلها أكثر وثوقيه وهذا يتحقق بفضل توفر البيانات داخل النظام حتى و إن كانت هناك بعض نقاط النظام خارجه عن العمل في وقت معين.

تعدد نسخ الملفات داخل النظام ساعد على وفرة هذه البيانات ،ولكن احتمالات أن تكون هناك نقاط خارجه عن النظام وفي حوزتها نسخ للملفات زاد في فرضية الحصول على بيانات غير متوافقة. وهذا ما جعلنا نقدم في هذا البحث استراتيجية تساعد في استمرارية العمل من خلال الجمع بين آلية التعديل في الزمن الحقيقي وكذلك آلية التعديل الآجل وتقنية صناديق البريد بما يساهم في الحصول على بيانات متوافقة داخل هذه النسخ.

٦- المراجع

1-M.Tamer Ozsu,Patrick Valduriez “Principles of Distributed Database Systems “ Pearson education Inc. 1999.

2-Taher Alrashahy “Model and Analysis for Data Allocation in Distributed Information System” Ph-D- thesis, 2001.

3-Ouri Wolfson, y.Huang “Competitive analysis of caching in distributed database”, IEEE Transaction on Parallel and distributed systems, April 1998 , PP 391-408.

٤-Bob Breton “Reliable Data Replication” Byte April 1998, PP 57-58.

٥-Deborah Russell “Oracle Distributed System “Charles Dye April 1999,PP 189-191.

6-Ajay M.Rmbhia " XML Distributed Systems Design first Printing " March 2002, PP221-269.

٧ - د.منيب قطيشات "قواعد بيانات" دار وائل للنشر الطبعة الثانية ٢٠٠٥- صفحة "٢٠٦/ ٢٠٧".