

Storage Fusion: DBMS における排他待ち問題解決支援機能の開発

西川 記史[†] 茂木 和彦[†]
[†](株)日立製作所 システム開発研究所

河村 信男[‡] 根本 直一[†]
[‡](株)日立製作所 ソフトウェア事業部

1. はじめに

デジタル世界に毎年追加される情報は、2006 年では 151 エクサバイト、2010 年には 988 エクサバイトに達すると見込まれる[1]。このデータ量の急増は人手によるデータ管理の限界を超えているという事実を直視する必要がある。また、大量のデータを管理するスキルを要する人材は極めて枯渇しており、管理を如何に容易化するか、今後のデータ処理システムを利用していく上で重要な課題となっている。

データ処理の主要なシステムとして、データベース管理システム(以下 DBMS と記す)が挙げられるが、その運用管理において、性能障害の解決に関する作業が全体の 4 割以上を占めることが示されている[2]。データ管理を容易化することで DBMS の性能障害の解決の容易化は最も重要な課題と考えることができる。この課題に取組むため、我々は性能問題の解決に関する作業の内訳の調査を実施した。この結果、上記作業の約 3 割が排他待ち問題の解決に充てられており、最も解決が難しくかつ有効な支援方法が存在しないことが明らかとなった。

これらの背景から、我々は DBMS における排他待ち問題の発生原因の特定を支援する排他待ち問題解決支援機能の開発を行った。本論文ではその内容について述べる。

2. 排他待ち問題解決支援機能

2.1 排他待ち問題の概要

排他待ちの発生原因としてまず考えられるのは、データベースの行などの排他対象となるリソース(排他対象リソース)へのアクセス集中による競合の発生である。しかし前述の調査の結果、(1) 排他待ち問題が顕在化した時点では既にその大元の原因となった処理が終了しており待ちの発生状況を追跡することが困難であること、(2)ある排他対象リソースで発生した待ちが、当該リソースにアクセスしない他のトランザクションの排他待ちを引き起こすケースがあり、排他待ち問題が検知された排他対象リソースに閉じた診断では排他待ちの大

元の原因となった処理を特定できないこと、(3) CPU やディスクなど排他リソース以外のリソース競合によるトランザクションレスポンスの増加によっても排他待ちが発生するなど、その発生原因が多面に渡ること、が排他待ち問題の診断を困難にしていることが分かった。

2.2 排他待ち問題解決支援機能の概要

上記のような排他待ち問題に対応するため、我々は、a) DBMS が実行する排他処理のトレースを収集する排他トレース収集機能、b) 収集した排他・SQL トレースを解析し、排他競合状況を可視化する排他競合状況可視化機能、及び c) 排他待ちの大元の原因となった SQL の処理及び所要時間の内訳を表示する SQL 所要時間内訳可視化機能を有する排他待ち問題解決支援機能を開発した(図 1)。以下、これらの概要を示す。

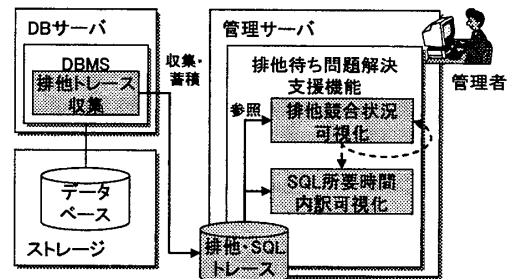


図 1. 排他待ち問題解決支援機能の概要

2.3 排他トレース収集機能

排他トレース収集機能は、排他待ち問題の診断に必要なデータを DBMS から収集する機能である。DBMS が行う排他確保・解放処理の頻度は非常に高い。このため我々は、排他待ちに関するトレースのみを収集することで排他トレース収集の負荷低減を図った。

2.4 排他競合状況可視化機能

複数のリソースに跨って発生する排他待ちには、トランザクションが排他リソースに対する排他を確保した後、当該トランザクションが他の排他リソースを確保しようとして待たされることが連鎖することにより発生する。この状況の調査のためには、排他リソースに対する排他を長時間保持していたトランザクションが他のトランザクション

により待たされていないか、待たされている場合はどのトランザクションのどの SQL が排他を長時間保持していたのかを明らかとすることが有効であると考えられる。

上記の診断を支援するため、我々は、排他リソースに対して長時間の排他を行っていたトランザクションの特定を支援する排他リソース競合状況可視化機能、及び排他待ちに直接的又は間接的に影響を与えた一つ又は複数の SQL の特定を支援するトランザクション動作状況可視化機能の2機能を用いて排他待ちの大元の原因となった SQL の特定を支援する技術を開発した(図2)。

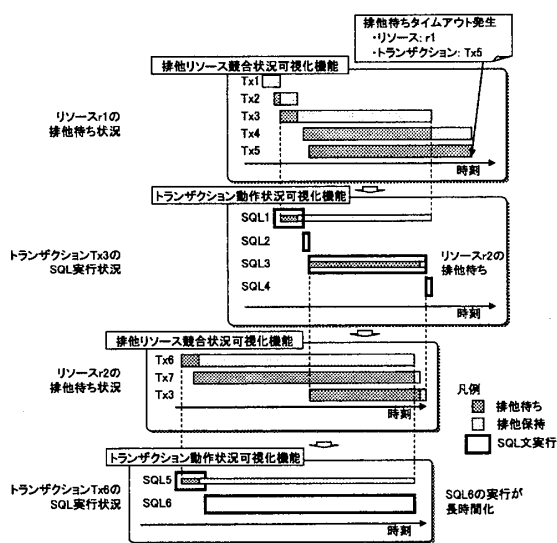


図2. 排他待ち状況の診断

2.5 SQL 所要時間内訳可視化機能

本機能は、排他待ちの大元の原因となった SQL の実行時間が長い原因を調査するために有用と考えられる情報を管理者に提示する。具体的には、当該 SQL の応答時間、CPU 使用時間、排他待ち時間、I/O 待ち時間、実 I/O 回数、SQL 文テキストを表示する。

3. 実験

3.1 排他待ち問題原因の特定支援

我々は、上記排他待ち問題解決支援機能の有効性を確認するための実験を実施した(図3)。本実験の結果、(1)従来排他待ちの原因として抽出されていた、排他を長時間保持していたトランザクション(図3左上B)が実際には他の排他リソースへのアクセスで排他待ちになっていたこと(図3②)、その原因はトランザクションCであり(図3③)、さらにトランザクションCが実行したSQLの実行時間が長いこと(図3④)が排他待ちの原因であることを明

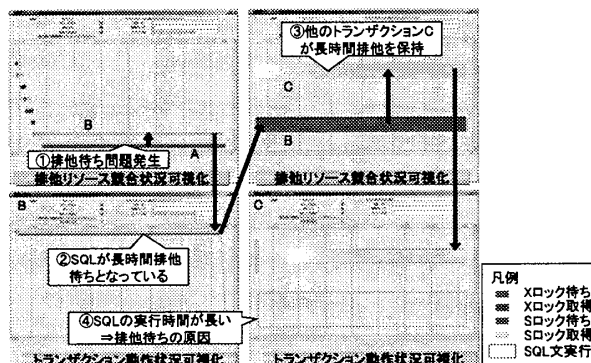


図3. 排他待ち問題の診断例

らかにすることができた。

3.2 排他トレース収集オーバーヘッド

排他トレース収集のオーバーヘッドを確認するため、TPC-C 相当のプログラムを用い、全ての排他トレースを取得する場合と提案方式の場合のトレース量を比較した。この結果、提案方式により、排他トレース量の大幅な削減が可能であることを確認した。

4. まとめ

DBMS の性能管理において最も解決が困難な排他待ち問題の解決を支援する技術を開発し、その有効性を確認した。その結果、これまで原因の特定がほとんど不可能であった、複数の排他リソース跨るに排他待ちの原因となった SQL の特定を行うことが可能となった。

謝辞

本研究に関してご指導いただきました東京大学の喜連川教授に深く感謝致します。

本内容に記載された技術には、文部科学省が実施するリーディングプロジェクト「e-Society 基盤ソフトウェアの総合開発」のストレージ・データベース融合技術(東大、日立)で技術開発された成果が反映されています。

参考文献

- [1] The Expanding Digital Universe, IDC White Paper, http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf, 2007
- [2] e-Society 基盤ソフトウェアの総合開発 先進的なストレージおよび Web 解析技術, http://cif.iis.u-tokyo.ac.jp/e-society/database/Kitsuregawa_t.html, 2004