

データベース同期ミドルウェアによる遠隔バックアップの活用手法の検討

細谷 柚子[†] 三島 健^{††} 小口 正人[†]

[†] お茶の水女子大学 〒112-8610 東京都文京区大塚 2-1-1

^{††} NTT ソフトウェアイノベーションセンター 〒180-8585 東京都武蔵野市緑町 3-9-11

E-mail: [†]yuzuko@ogl.is.ocha.ac.jp, ^{††}mishima.takeshi@lab.ntt.co.jp, ^{†††}oguchi@computer.org

あらまし 近年、災害対策として遠隔バックアップの必要性が問われている。ローカルのリソースだけでシステムを運用していると、大きな災害時にバックアップも失ってしまうからである。日本のような地震国では特に、遠方にバックアップを置いておくことがますます重要になっている。本研究は上記バックアップの対象として、企業でのサービス提供において大規模災害に対する耐障害性保証が特に求められるデータベース・バックアップにフォーカスする。その中でも Web サービスへの適用性が高いクラスタ DB(Pangea) を前提として、サービスのためのトランザクション処理に影響を及ぼさないリモートバックアップのためのミドルウェア Pangea++について検討した。そのミドルウェアを検証した結果、クライアントからのトランザクション処理においては、スループット性能に殆ど影響を及ぼさないとともにバックアップを最新に更新できる方式であることを確認した。

キーワード 遠隔バックアップ, データベース同期, 分散処理

A Study of Remote Backup Utilization Technique based Database Replication Middleware

Yuzuko HOSOYA[†], Takeshi MISHIMA^{††}, and Masato OGUCHI[†]

[†] Ochanomizu University Otsuka 2-1-1, Bunkyo-Ku, Tokyo 112-8610 Japan

^{††} NTT Software Innovation Center Midoricho 3-9-11, Musashino-Shi, Tokyo 180-8585 Japan

E-mail: [†]yuzuko@ogl.is.ocha.ac.jp, ^{††}mishima.takeshi@lab.ntt.co.jp, ^{†††}oguchi@computer.org

1. はじめに

近年ビジネスフィールドにおけるバックアップの需要が高まっている。TechTarget ジャパンによるデータ保護・管理対策に関するアンケート調査 [1] によると、全体の約 90 % がバックアップを実施していると答えたが、その一方で 3 分の 1 以上が現状に何らかの不満を抱いていた。その理由として、バックアップのパフォーマンスや、維持にかかるコストなどが挙げられ、現状のバックアップに対する不安が顕著に表れていた。従ってバックアップのパフォーマンス向上はまだまだ課題が残っていると言える。さらに、2011 年 3 月の東日本大震災以降から遠隔バックアップの必要性が急速に求められている。ローカルのリソースだけでシステムを運用していたら、大規模災害時にはバックアップごと失ってしまうからである。情報社会において、バックアップを失うことは、システムの停止に直接つながり、企業にも顧客にも大きな損害を及ぼしてしまうことが懸念される。ゆえに日本のような地震国では、遠方にバックアップを置いておくことがますます重要になっている。

本研究では、特に、データベースのバックアップに注目した。データベースは企業の基幹を構成しており、バックアップをとるべき重要度の高いシステムと考えるためである。

LAN 環境を前提としてデータベースを同期する Pangea [2] という既存のミドルウェアがある。これを拡張した新たなバックアップを確立するミドルウェア Pangea++ を提案し、TPC-W ベンチマーク [3] を用いて評価実験を行った。評価実験より、提案する手法の課題抽出と、更なる性能向上を目指した解決手法について検討を行う。

2. 先行研究:Pangea

LAN 環境を前提として、データベースの同期を行うミドルウェアの Pangea がある。Pangea は、サーバの 1 台を Leader、その他は Follower としており、クライアントから Pangea のミドルウェアを介してサーバにアクセスすることで同期をとる。照会処理は 1 台のサーバで実行され、更新処理は全てのサーバで実行される。更新処理の場合は Leader に対して更新をし

た後に、Follower に対しても同様に処理を行う。データベース (DBMS) を修正することなくサーバを増やすことで性能を向上させることが可能である。

3. Pangea の遠隔バックアップ確立

まず、Pangea を修正せずにそのまま用いて遠隔バックアップを確立させる予備評価を行った。このとき、Follower を遠方に置くことで遠隔バックアップとした。データベースの実行処理を分担する Follower のデータベースは同期されていることから、これがバックアップとして機能する。サーバマシンは Leader、Follower 用に 1 台ずつ用意した。マシンのスペックを表 1 に示す。バックアップは海外にあることを想定し、Dummynet を使用して RTT256ms の遅延を挿入した。データベースは PostgreSQL9.2.6 [4] を使用し、すべてのマシンに配置させた。Web サーバとアプリケーションサーバには Tomcat6.0.37 [5] を用いた。

性能評価は TPC-W ベンチマークを使用した。TPC-W は仮想的なブラウザ (以下 EB とする) が、データベースにトランザクションを発行する。TPC-W には 3 種類のワークロードがあり、それらの違いは表 2 に示すように照会処理と更新処理の割合が異なる。性能評価指標は、スループット (1 秒あたりの Web 画面表示:(WIPS)) とレスポンス時間 (1 画面データの転送時間:(秒)) とした。実験環境を図 1 に示す。

表 1 マシンのスペック

OS	Ubuntu14.04
CPU	Intel(R) Xeon(R) CPU 1.60GHz (4cores) / Intel(R) Xeon(R) CPU 1.60GHz (1core)
Memory	2GByte / 4GByte

表 2 ワークロード

ワークロード	Read-only	Update
Browsing mix	95 %	5 %
Shopping mix	80 %	20 %
Ordering mix	50 %	50 %

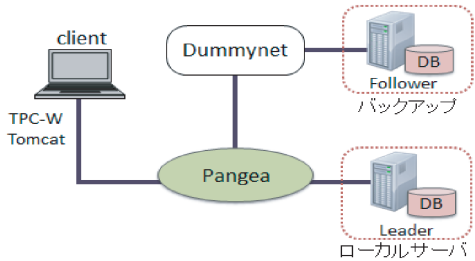


図 1 Pangea の実験環境

TPC-W の 3 種のワークロードそれぞれで実験を行った。Follower をローカルに配置した場合と、遠方に配置した場合とで性能の比較をした。結果を図 2, 3, 4 に示す。

棒グラフはスループット、線グラフはレスポンス時間をして

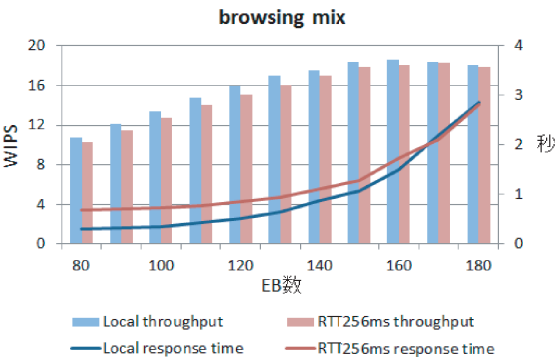


図 2 Pangea の browsing mix の性能

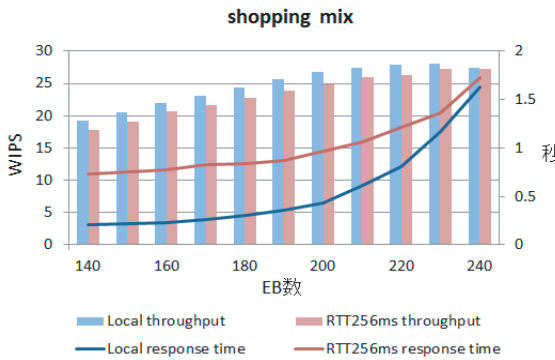


図 3 Pangea の shopping mix の性能

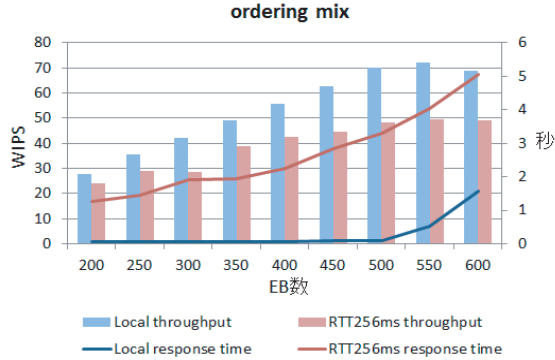


図 4 Pangea の ordering mix の性能

いる。横軸は EB 数である。Follower がローカルにある場合の Pangea の最大スループット値は、browsing mix では、EB が 160 のとき 18.61WIPS、shopping mix では、EB が 230 のとき 27.98WIPS、ordering mix では、EB が 550 のとき 72.15WIPS となり Follower が遠方にある場合の Pangea の最大スループット値は、browsing mix では、EB が 170 のとき 18.24WIPS、shopping mix では、EB が 240 のとき 27.28WIPS、ordering mix では、EB が 550 のとき 49.37WIPS となった。

Follower を遠方に置くことによるスループットの低下率は、browsing mix では最大で 5.6 %、最小で 0.71 %、shopping mix では最大で 7.22 %、最小で 0.76 % ordering mix では最大で 31.57 %で、最小で 14.25 %あった。レスポンス時間は、browsing mix では最大で 0.38 秒の増加、最小で 0.05 秒の低

下 shopping mix では最大で 0.57 秒，最小で 0.1 秒の増加，ordering mix では最大で 3.54 秒，最小で 1.25 秒の増加がみられた．このように，全てのワークロードで性能が低下するのは，遠方にある Follower でクエリを処理するためである．また，ordering mix で最も性能が悪化した理由は，遠方にある Follower で実行するクエリが最も多いからである．

この結果から，遠隔バックアップ確立のパフォーマンスを向上させることを本稿の課題とし，Pangea を拡張させた，新たなバックアップ確立ミドルウェア Pangea++ を提案する．遠隔バックアップにおいては，データベースの照会処理をする必要はなく，更新処理のみを行えばよいため，Pangea++ ではこの点を改良する．

4. 提案手法

クライアントからの処理を分担するローカルのサーバを用意する．既存研究と同様に，1 台を Leader，その他を Follower とする．それらとは別に，データベースの実行処理を分担しない遠隔バックアップ用のサーバを用意する．バックアップサーバは，Pangea を介して更新処理のみ行われるように拡張した．これを Pangea++ と呼ぶ．

図 5 のように，ローカルサーバの処理をする複数の worker スレッドとバックアップの処理をする synchronizer スレッドを 1 つ用意する．worker スレッドはクライアントからクエリが来ると，ローカルサーバに対して処理を行うとともに，送られてきたクエリに関して，更新クエリのみをトランザクションごとに保管する．これを writeset と呼び commit 順に保管する．worker は commit を契機にして，synchronizer スレッドを呼び起こす．synchronizer スレッドは，worker が保管しておいたクエリを先頭から順に取り出し，バックアップサーバに対して処理を実行する．提案手法により，バックアップ処理を worker から切り離すことにより通常処理の性能低下を抑えると期待できる．

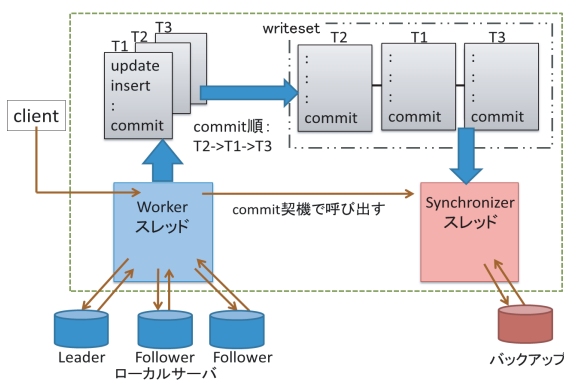


図 5 提案手法

5. 評価実験

5.1 実験環境

提案手法の性能評価実験を行った．ローカルサーバ用にマシン 2 台，バックアップサーバ用にマシン 1 台を用いた．バ

ックアップは海外にあることを想定し，Dummysnet を使用して RTT256ms の遅延を挿入した．マシンや，データベース，Web サーバ，アプリケーションサーバ，ベンチマークは 3 節と同様のものを用いた．実験環境を図 6 に示す．

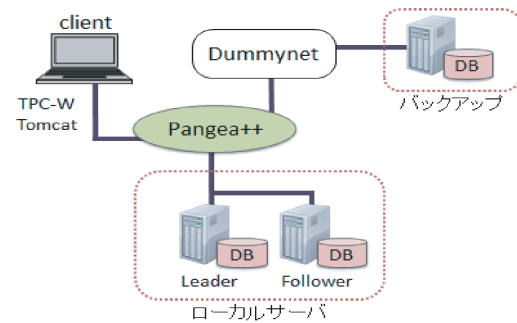


図 6 Pangea++の実験環境

5.2 Pangea++の性能

TPC-W の 3 種のワークロードそれぞれで実験を行った．3 節の Pangea の遠隔バックアップ確立手法と比較した結果を図 7，8，9 に示す．

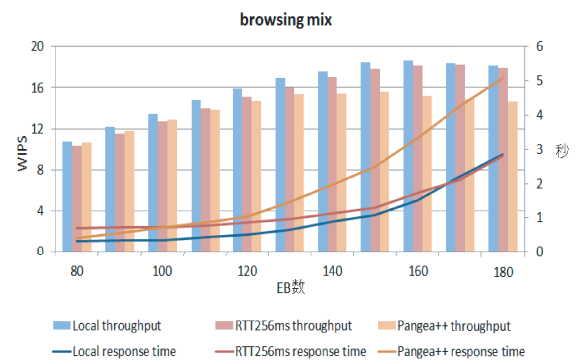


図 7 browsing mix の Pangea と Pangea++ との比較

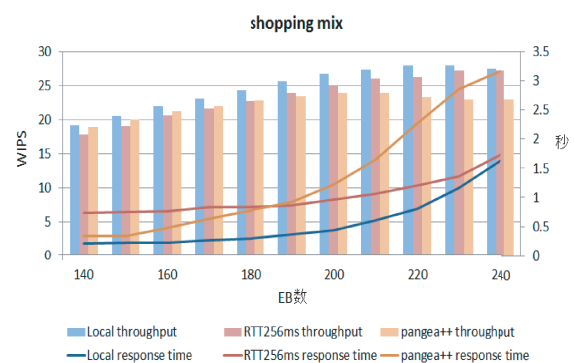


図 8 shopping mix の Pangea と Pangea++ との比較

棒グラフがスループット，線グラフがレスポンス時間を示している．横軸は EB 数である．Pangea の Follower をローカルに配置した場合を Local，Follower を遠方に配置した場合を RTT256ms，提案手法を Pangea++ としている．Pangea++ の最大スループット値は，browsing mix では，EB が 150

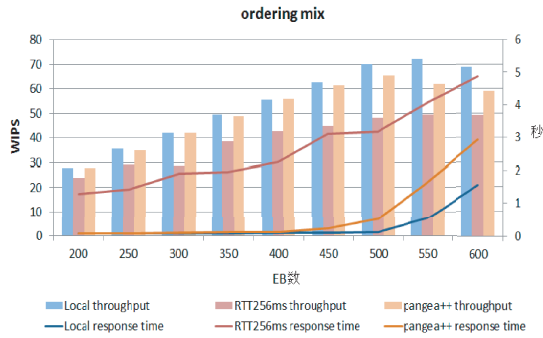


図 9 ordering mix の Pangea と Pangea++ との比較

のとき 15.59WIPS, shopping mix では, EB が 200 のとき 24.02WIPS, ordering mix では, EB が 500 のとき 65.48WIPS であった. RTT256ms と Pangea++ のスループットの性能差は, browsing mix では最大で 22.33 % の低下, 最小で 1.23 % の増加, shopping mix では最大で 15.79 % の低下, 最小で 0.4 % の増加, ordering mix では最大で 31.63 %, 最小で 14.95 % の増加があり, レスポンス時間は browsing mix では最大で 2.26 秒の増加, 最小で 0.003 秒の減少, shopping mix では最大で 1.49 秒, 最小で 0.05 秒増加, ordering mix では最大で 2.89 秒, 最小で 1.25 秒の減少が見られた.

browsing mix と shopping mix は更新クエリよりも参照クエリが多く, 遅延の影響をあまり受けなかったことで, RTT256ms は Local よりもわずかな性能低下である. これに対し Pangea++ では, writeset を保管することがオーバーヘッドになり, RTT256ms よりもさらに性能が低くなった. 一方で ordering mix では, writeset を保管するオーバーヘッドから Local よりも Pangea++ の性能低下が見られたが, Pangea++ ではバックアップはトランザクション処理を行わないため, RTT256ms と比較すると性能が向上した. この結果から, 参照クエリの多い場合には Pangea の遠隔バックアップが有効であり, 更新クエリの多い場合には Pangea++ が有効であることがわかった.

5.3 バックアップの処理性能

次に, 負荷を変動させて writeset のキューの長さを測定し, バックアップの処理性能を調査した. 負荷は TPC-W の EB 数を変化させることで調整した (表 3). 低負荷の調査結果を図 10, 11, 12 に, 高負荷の調査結果を図 13, 14, 15 に示す.

表 3 測定した負荷

負荷	EB 数
高負荷	最大スループット時の EB 数
低負荷	10 に固定

縦軸がキューの長さ, 横軸が commit 数を示している. 低負荷の場合は, キューの長さが 0 になることがあるため, 処理がスムーズに行われていることがわかる. 高負荷の場合は, 低負荷の場合と比較すると, 全てのワークロードで writeset に溜まるキューが増えている. browsing mix では辛うじてキューの長さ

バックアップ処理性能(browsing mix)

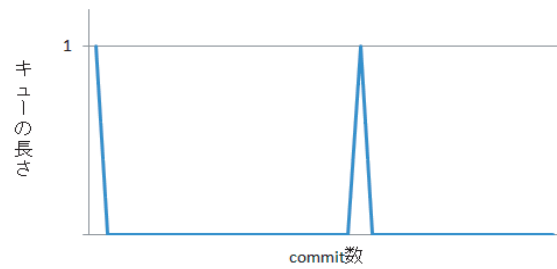


図 10 低負荷時の browsing mix のキュー長

バックアップ処理性能(shopping mix)

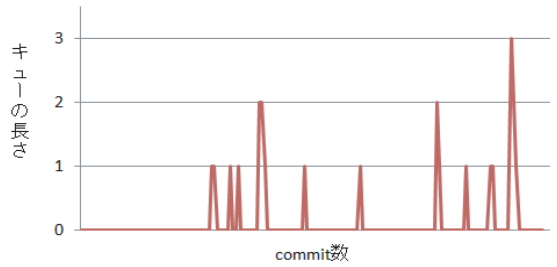


図 11 低負荷時の shopping mix のキュー長

バックアップ処理能力(ordering mix)

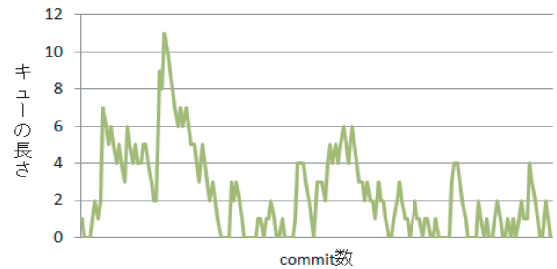


図 12 低負荷時の ordering mix のキュー長

バックアップ処理能力(browsing mix)

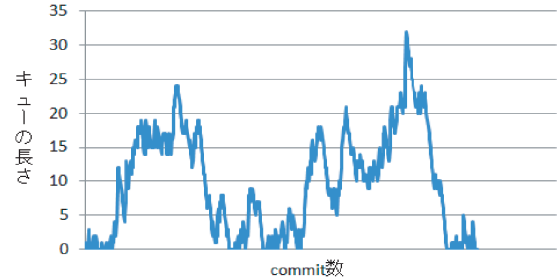


図 13 高負荷時の browsing mix のキュー長

が 0 になることがあるが, 一方で shopping mix と ordering mix は右上がりのグラフになっていることから, writeset のキューが溜まってしまい輻輳していることがわかった. EB 数を増やした場合であっても, バックアップはキューを溜めずに処理をこなすことが望ましく, 今後, 改善すべき課題であると考える.

6. まとめと今後の課題

データベース同期ミドルウェア Pangea を拡張したバックアップ



図 14 高負荷時の shopping mix のキュー長

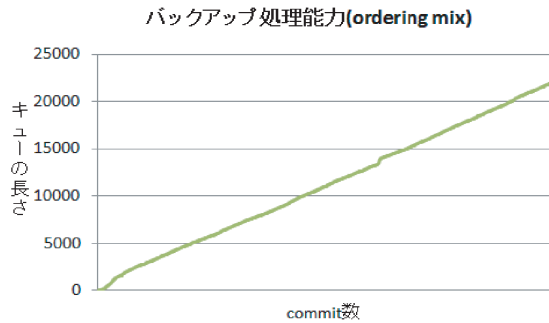


図 15 高負荷時の ordering mix のキュー長

ブミドルウェア Pangea++の提案と評価を行った．Pangea のバックアップ確立手法と Pangea++の性能比較を行い，参照クエリの多いワークロードでは Pangea を利用したバックアップが望ましく，更新クエリの多い場合では Pangea++が望ましいと考える．さらに，バックアップの処理性能を調査するため，writeset のキューの長さを測定した．TPC-W の EB 数を増やし，負荷を増大させた場合に同期のための writeset が溜まってしまう課題があることがわかった．

今後の課題としては，負荷を増大させた場合においても，キューを溜めずに処理をこなせるように，提案手法の改善を行いたい．また，バックアップサーバにも負荷分散させる手法についても検討し，バックアップ確立の更なるパフォーマンス向上を目指したい．

文 献

- [1] データ保護・管理対策に関するアンケートリポート
<http://techtarget.itmedia.co.jp/it/news/1109/27/news03.html>
- [2] T.Mishima and H.Nakamura,"Pangea:An Eager Database Replication Middleware guaranteeing Snapshot Isolation without Modification of Database Servers", Proc.VLDB2009, pp.1066-1077, August 2009. PVLDB2009.
- [3] TPC-W <http://www.tpc.org/tpcw>
- [4] PostgreSQL <https://www.postgresql.jp/>
- [5] Tomcat <http://tomcat.apache.org/>