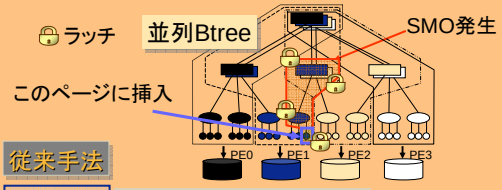
平成18年度の成果

平成17度までの成果は 第3回シンポジウム 講演要旨をご覧下さい

スケーラビリティに優れる並列B⁺-tree並行制御手法

並列Btree 無共有並列計算機上の並列DBに適した効率的なインデックス構造

並行性制御 Btree構造更新時の一貫性を保証 システム処理性能に影響大



従来手法

- **ARIES/IM** PE間同期オーバーヘッド大
- **INC-OPT** PE間同期オーバーヘッド小 リスタート回数大
- **MARK-OPT** マークを用いた先読みによるリスタート削減

ラッチカップリングを用いたリクエスト移譲 SMO関連全ノードでのラッチ獲得後の更新

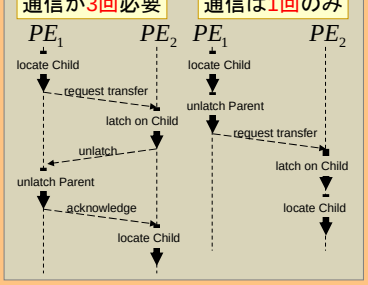
→ 分散環境における検索更新コスト高

LCFB

ラッチカップリングの省略

リクエスト移譲コスト削減

非ラッチカップリングにより発生する問題を解決し 正しいBtreeの並行性制御を実現



更新コストの低い B-linkの利用

B-linkを部分的に適用

LCFBとB-linkを統合

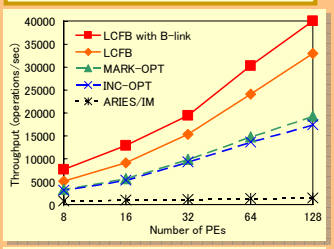
更新コスト削減

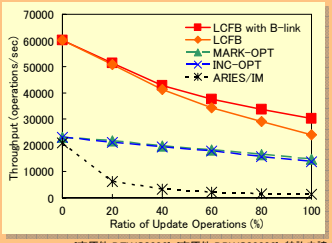
検索コスト削減

→ 効率的な検索と更新の実現

従来手法の性能を常に改善

大規模高更新環境では大幅改善





[吉原他 DEWS2006], [吉原他 DBWS20006], 特許申請中

スケーラビリティに優れる更新ログコミットプロトコル

背景

自律ディスクのような分散ストレージ
データの正当性保証のための分散コミットプロトコル(CP)が必須

自律ディスク上で効率的なCPの要求
2 Phase CP等の既存CPでは通信コストやディスクI/Oコストが高い

目的

自律ディスク特性を考慮したCPの効率化

アプローチ

- ロギングにおけるディスクI/Oブロッキングの除去
- コミット処理における投票フェーズの除去
- Primary-Backup方式の補助による障害復旧

非同期neighbor-WAL

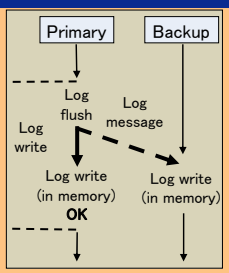
Neighbor-WALを利用
自PEではなく他PE主記憶へのログ書込
ディスクI/Oログオーバヘッドを軽減

ログメッセージの非同期化
複製PE間同期オーバヘッドの削減
複製PE間の同期は最終決議フェーズ

1.5フェーズコミットプロトコル

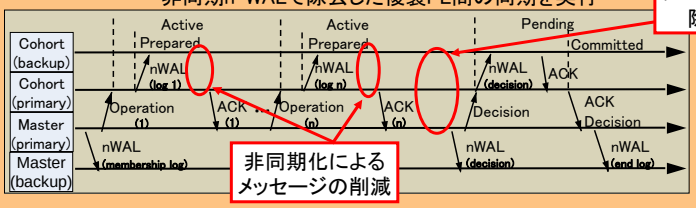
投票フェーズの除去
投票メッセージは各操作の応答に暗黙的に包含
高速なコミットプロセスを実現

決議フェーズでコホートバックアップ待ち発生
非同期n-WALで除去した複製PE間の同期を実行

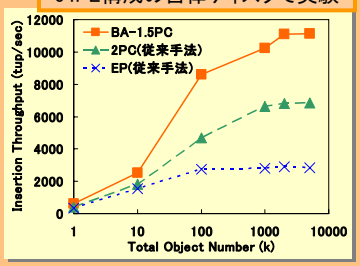


障害復旧戦略

プロトコルの各フェーズにおける障害に対応
プライマリコホートの障害
バックアップコホートで対応
決議フェーズにおけるマスタ障害
コホートの処理をブロックせず



64PE構成の自律ディスクで実験



大規模環境で従来手法の性能を常に改善

提案手法のオーバヘッド削減効果を確認

[Ouyang et al. PRDC20006], 特許申請中

スケーラビリティに優れる自律管理ルール高速処理手法

背景と目的

ECAルールによるルール記述

アクティブDBで利用されるECA (Event-Condition-Action)によるルール記述・管理

柔軟かつ細粒度のコンテンツ管理

コンテンツにメタデータの一部としてルールを定義
→ ルール数がコンテンツ数と共に爆発的に増加

大量のルールを高速に条件評価する必要性

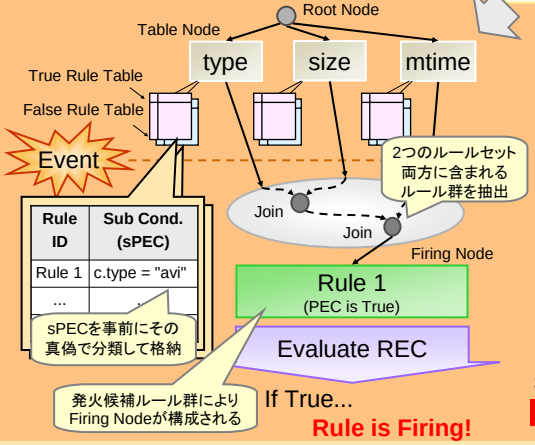
アプローチ

ルールの発火条件のうち事前評価可能な部分の評価をイベント発生前に実行して発火候補ルールを絞り込む

ルールの発火条件を事前実行の可否により分割:
PEC :イベント発生前に評価可能な条件
sPEC :PECを単一条件式ごとに分割したもの
REC :イベント発生時に評価すべき条件

提案手法

従来は対象データ(タプル)を流していた
弁別ネットワークに対象ルールを流し,
イベント発生時に事前評価済みPECで絞込み

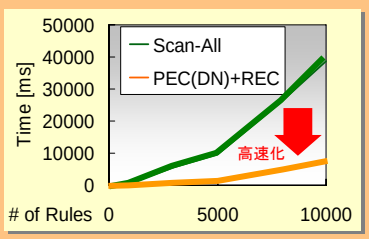


Event Cond. 定期的な更新チェックイベント発生時 PEC
AVIファイル sPEC
かつ 500 MB 以下
かつ 2005年12月9日17時以前に更新され
かつ 期限が切れていないコンテンツに対し REC
Action 低速アーカイブへ移動

コンテンツ管理ルールの例

実験結果

ルール処理時間の比較



提案手法は全走査手法と比較して大幅に処理時間を短縮することが可能

[Ohta et al. SWOD2006]