

Tokyo Tech

集積したデータを
早く正確に
解析するために

2021年10月11日 剣ユーザ会

横田 治夫
東京工業大学 情報理工学院

やりたかったこと、諦めてきたこと

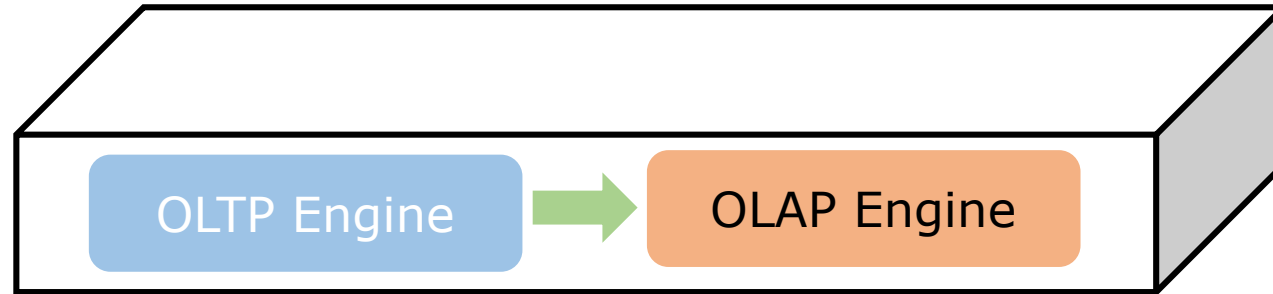
- やりたかったこと（大昔から多くの研究者）
 - 集積したデータを早く正確に解析する（究極の課題）
 - データの集積は OLTP (On Line Transaction Processing)
 - データの解析は OLAP (On Line Analytical Processing)
- 皆さん、これまで諦めていませんでしたか
 - 厳密な正確さを求めるなら OLTP が全部終わってから
 - 週末、夜に処理
 - 早さを求めるなら OLAP の結果の正確さは犠牲に
 - Dirty Read、Non-Repetable Read、Phantom Read
 - Snapshot Isolation の Skew Anomaly (含む Read Only Anomaly)
- 昨年のユーザ会では Anomaly の話が中心
 - 今年はシステムの前提についてお話してみたい
 - Anomaly については、昨年の資料をご参照下さい

やっていること、やろうとしていること

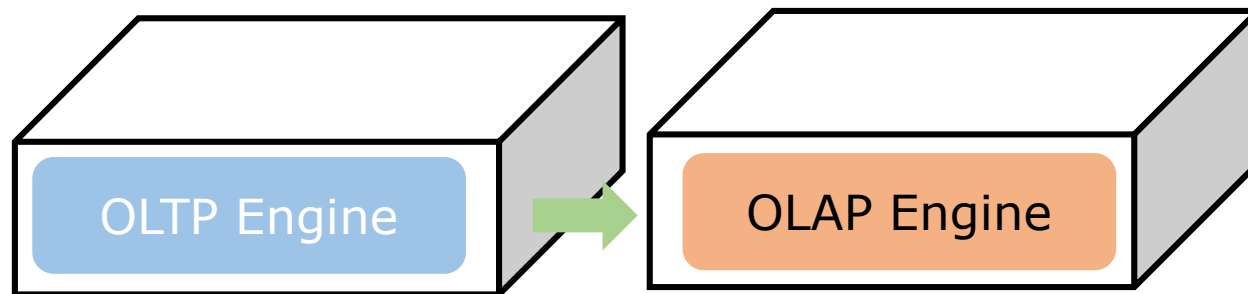
- 正確にできるだけ早く OLTP の更新内容を OLAP に適用
 - 現在、Anomaly なしの手法を実装、実験結果を評価し、論文執筆中
 - PostgreSQL での実装を公開予定
 - 場合によっては Tsurugi に含まれる可能性もあり
- その際、適用が早いだけでなく性能的にも速く
 - OLAP に影響されない OLTP 性能
 - OLAP によって OLTP は待たない、アボートしない
 - 限られた資源の中で OLTP の負荷の変動に対応 (動的資源配分)
- 異なるシステム構成を想定
 - A) マルチコアのシングルノードを使う構成
 - B) OLTP と OLAP で異なるノードを使う構成
 - C) ラック内の複数ノード構成での負荷分散

想定するシステム構成

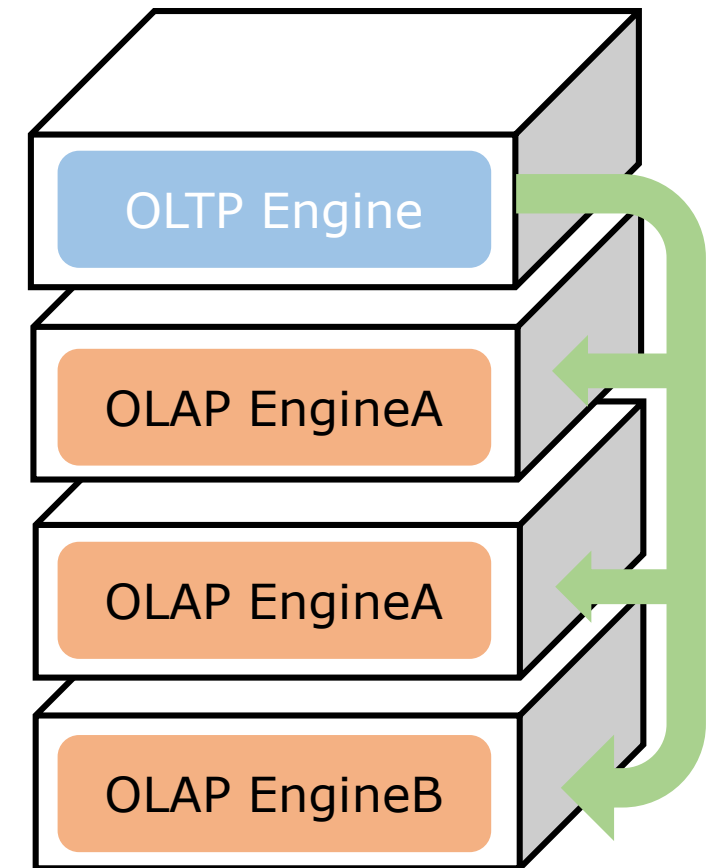
A) シングルノード (マルチコア) 構成



B) マルチノード構成

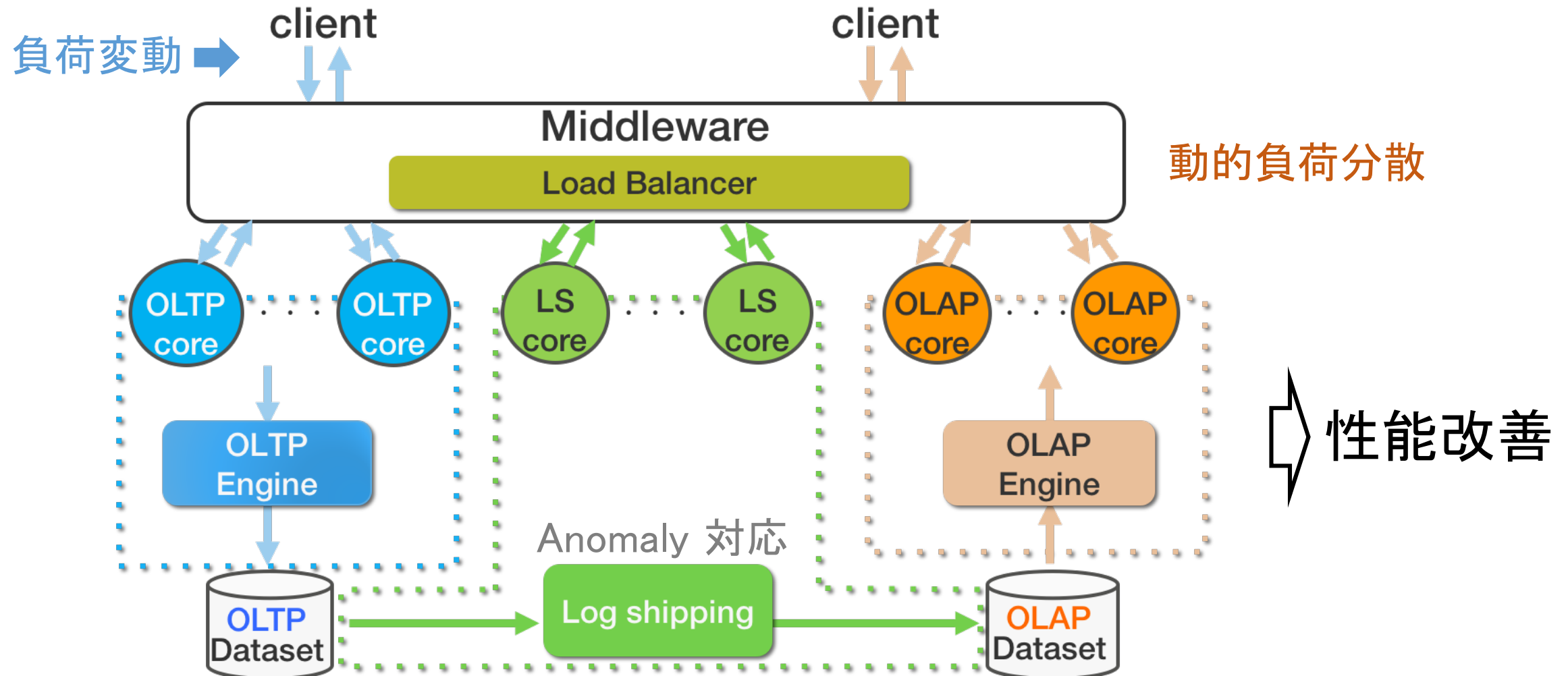


C) ラック構成

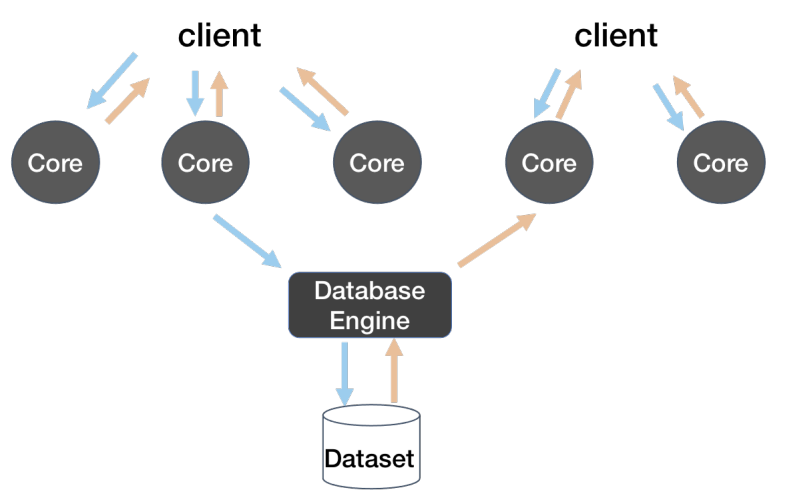


更新適用時の性能向上のための動的資源配分

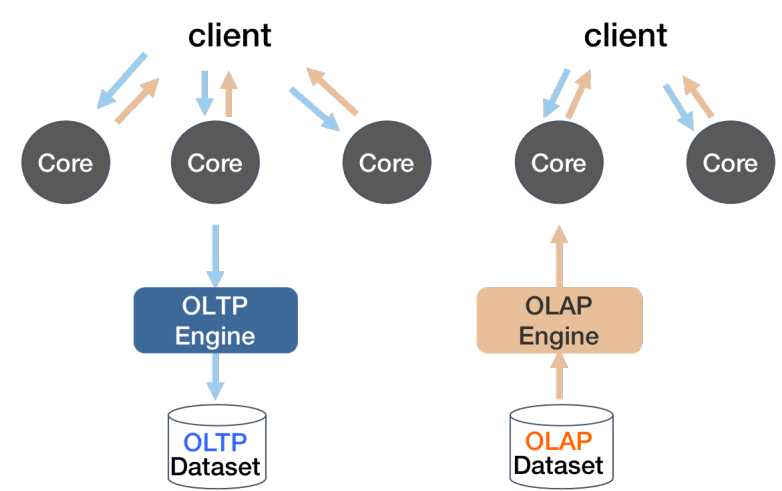
- 高優先順位OLTPの負荷の変動に対応して資源を調整



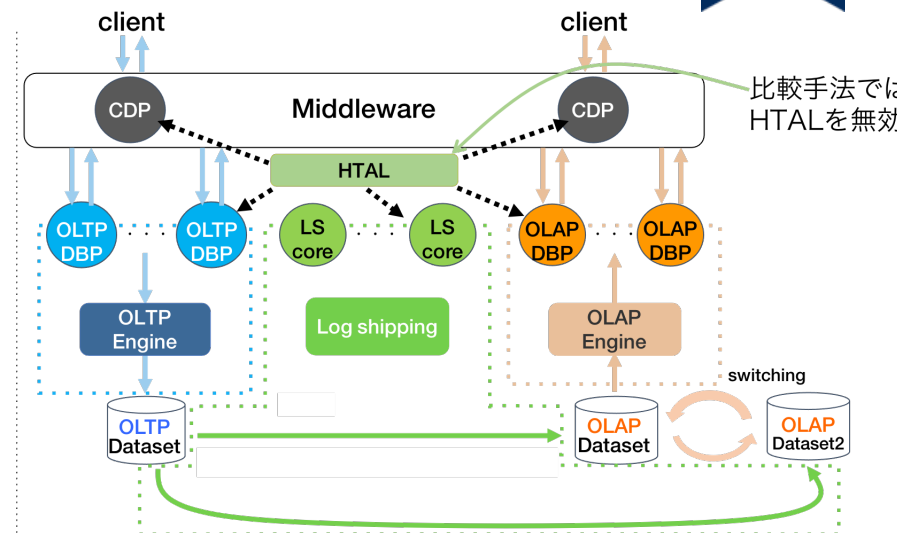
比較実験 [DEIM2021]



Baseline (Unified)



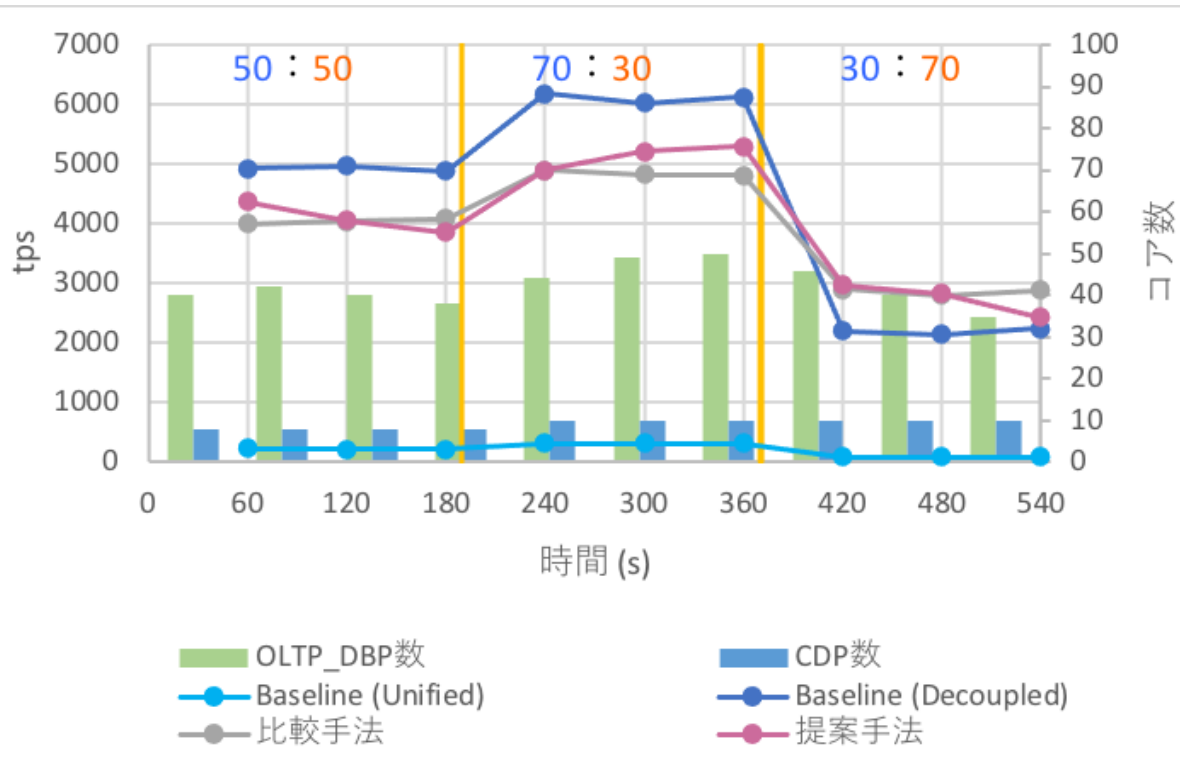
Baseline (Decoupled)



提案手法

	HTAL	コアアサイン	OLAPデータ切り替え	Log shipping	OLAPデータ分離	OLTP Isolation level	OLAP Isolation level
Baseline (Unified)	×	×	×	×	×	SERIALIZABLE	SERIALIZABLE
Baseline (Decoupled)	×	×	×	×	○	SERIALIZABLE	REPEATABLE READ
比較手法	×	○ (静的)	○	○	○	SERIALIZABLE	REPEATABLE READ
提案手法	○	○ (動的)	○	○	○	SERIALIZABLE	REPEATABLE READ

	サーバ	クライアント
プロセッサ	Intel Xeon Platinum 8260M	Intel Xeon E7-4860
ソケット数	2	4
論理コア数	96	80
L1キャッシュ	32 KB+32 KB	32 KB+32 KB
L2キャッシュ	1024 KB	256 KB
メモリ	384 GB	32 GB
二次記憶装置	OLTP : 1.5 TB(NVDIMM) OLAP : 880 GB(SSD)	200 GB(SSD)
OS	Ubuntu 18.04.5	Ubuntu 16.04.3
PostgreSQL	12.4	12.4



● コア数配分

- 50 : 50 OLTP_DBPが増減
- 70 : 30(OLTPヘビー) OLTP_DBPが増加
- 30 : 70(OLAPヘビー) OLTP_DBPが減少

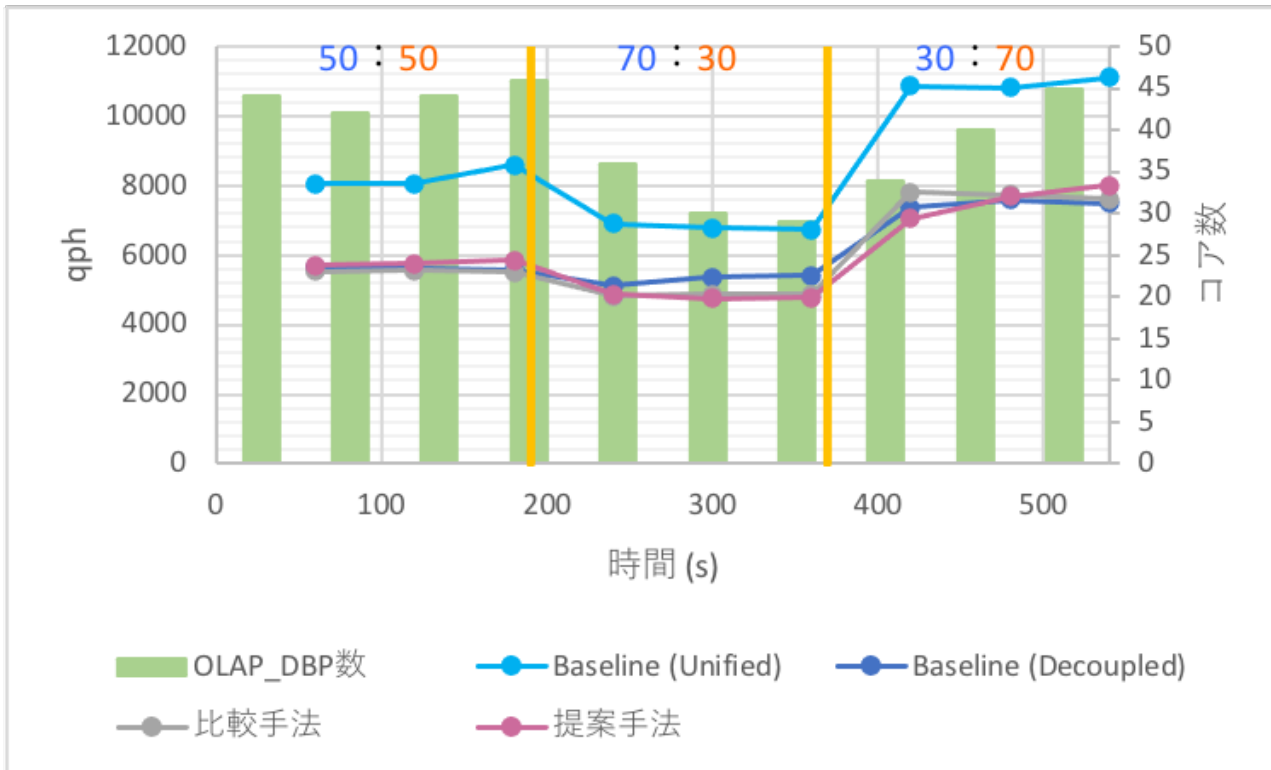
- OLTP_DBP減少時以外,比較手法より優れた性能 : 10.1%向上

● Baseline (Decoupled)との比較

- OLTPヘビー時にはBaselineに劣る性能
 - ✓ CDPやLSによってOLTP処理コアが減少するため
- OLAPヘビー時の性能では34.9%向上
 - ✓ コアアサインによってOLAP実行に伴う性能劣化が小さいため

- Baseline (Unified)はOLAPによりABORT率が増加
- 比較手法・提案手法は同一rangeにアクセスするDBPが指定されるため、Baseline (Decoupled)よりABORT率が低い
- 提案手法はHTALによってCDP・OLTP_DBP個数変化時にABORTを返すため比較手法よりABORT率が高い

	Baseline (Unified)	Baseline (Decoupled)	比較手法	提案手法
ABORT率	10.5%	8.2%	5.4%	6.4%
実行Tx数	98,958	2,104,867	1,928,263	2,009,562
ABORT数	10,444	168,389	104,126	128,611

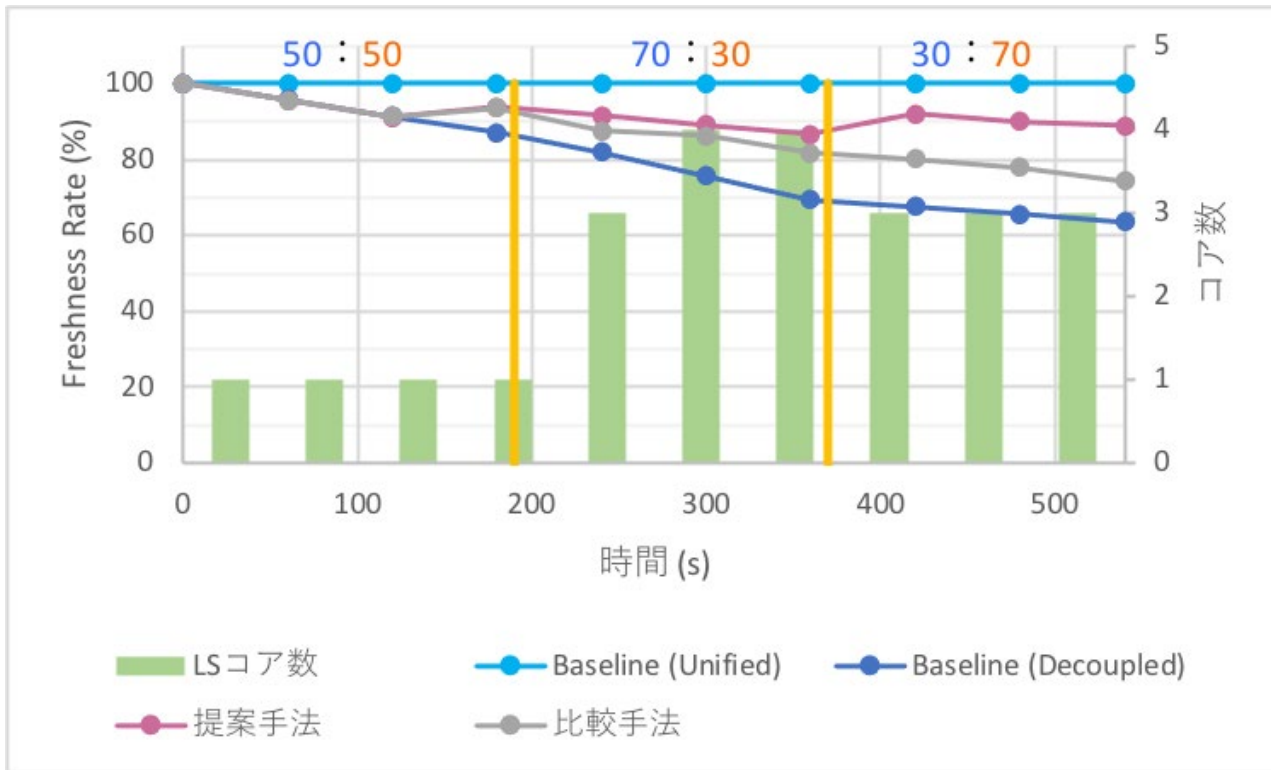


● コア数配分

- OLAPへビー時にはLSに多くコアアサインされ、Freshnessが優先

● OLAP処理性能

- OLAP_DBP数に応じて比較手法と性能差が生じる
- Freshnessが優先され、OLAP処理性能はBaselineや比較手法より性能低下



● OLAPデータの切り替え

- 比較手法：120秒おきに切り替え
- 提案手法：HTALによってFreshness低下時に切り替え

● コア数配分

- OLAPへビー時にもLSコアに多くアサイン

● LSコアの増加により比較手法より優れたFreshness

- 最大時19.9%高いFreshness Rate

- 目標

- 正確にできるだけ早く OLTP の更新内容を OLAP に適用
 - 適用が早いだけでなく性能的にも速く

- Anomaly なし適用手法の PostgreSQL 実装は評価中

- 最終的には公開を予定
 - 場合によっては Tsurugi に含まれる可能性もあり

- 様々な構成を想定

- 限られた資源の中で OLTP の負荷の変動に対応
 - 負荷分散手法も Tsurugi に適用可能



Tokyo Tech

ご清聴ありがとうございました

