

# 大規模データ連携処理の実現に向けた性能検討

細田 聖人<sup>†</sup> 楓 仁志<sup>†</sup> 石井 篤<sup>†</sup> 高山 茂伸<sup>†</sup>

三菱電機 情報技術総合研究所<sup>†</sup>

## 1.はじめに

従来、システム間でのデータ連携は、連携ごと個別のバッチプログラム等で構成されており、データ連携の規模増加に伴い連携の変更や管理が複雑になる、またデータ処理性能が劣化する、といった課題がある。筆者らは、複数システム間のデータ連携に対し、データ層での連携を集約するハブ型構成のデータ連携システムを提案しており、システムからの連携処理の独立、データ連携の基盤化による連携の管理と実行の一元化を実現する[1]。また、複数データ連携実行時に、データ受信・変換の共通化可能な処理を集約し、規模増加に対し性能劣化を抑止する方式を検討している[2]。本書では、連携の性能測定を実施し、上記方式の妥当性を確認する。

## 2.データ連携システムの構成

データ連携の大規模化に伴い、データ連携システムには、システム停止時のデータ復旧を実現する等の高信頼性が求められる。また、連携元システムでのデータ更新を即時に検知・抽出するCDC(Change Data Capture)技術の普及により、従来のバッチ連携だけでなくリアルタイムデータ連携を対象とする必要がある。上記の要求事項を考慮し、本書では図1のデータ連携構成を検討対象とする。データ連携の構成は、連携元システム・データ連携システム・連携先システムから成る。データ連携システムは受信・変換・反映の3つの機能を持つ。これらは順に、連携元システムからの更新データを受信する機能、連携先システムのデータ形式に変換する機能、連携先システムにデータ反映する機能である。また本構成では、受信機能と反映機能にてデータ復旧のためのジャーナルデータ書き込み処理を実施する。またリアルタイムデータ連携の高速化のため、データ連携処理は常駐プロセスにより実現される。

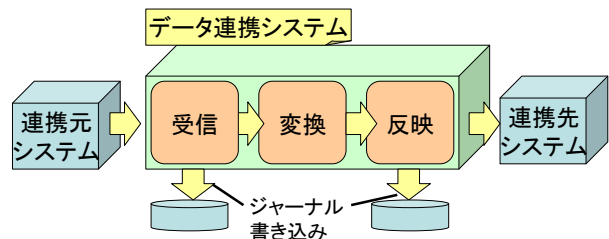


図1 データ連携の構成

## 3.性能測定

本性能測定においては、データ連携システムへのデータ到着から連携先へのデータ反映までを計測、データ連携システムにおける各機能の処理時間を測定した。リアルタイム連携処理からバッチ処理によるデータ連携まで想定し、1トランザクションのレコード数を1~300と変更し計測した。1レコード長は512byteである。なお本測定での変換処理は、5%のカラムに対する文字列変換処理である。性能測定の結果を図2に示す。

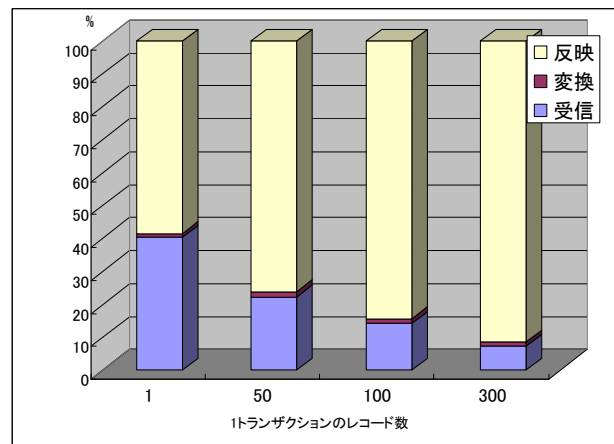


図2 性能測定結果

性能測定の結果より、1トランザクションあたり1レコードの場合は、受信処理は全体の40.3%、変換処理は全体の1.2%、反映処理は58.5%であった。また1トランザクションあたり300レコードの場合は、受信処理は全体の7.2%、変換処理は全体の1.3%、反映処理は91.5%であった。

トランザクションあたりのレコード数増加に伴い反映処理時間の割合が増加している要因は、受

A Basic Study of Performance Characteristic for Large Scale Data Integration.

<sup>†</sup>Kiyoto Hosoda, Satoshi Kaede, Atsushi Ishii, Shigenobu Takayama, Information Technology R&D Center, Mitsubishi Electric Corporation.

信機能と反映機能でのジャーナルデータ書込み処理時間の増加に比べ、連携先データベースに対する反映機能の SQL 処理時間の増加が大きいためであった。

## 4. 考察

### 4.1 連携増加時の集約効果

図 2 に示す各機能の処理時間から、連携増加時の、処理集約による効果について考察する。連携先が増加した場合を対象に、個別に連携を実施した場合と、処理集約した連携を実施した場合の連携処理を図 3 に示す。連携元での変更を 2 つの連携先に反映する場合は、受信処理と変換処理の共通化が可能である。

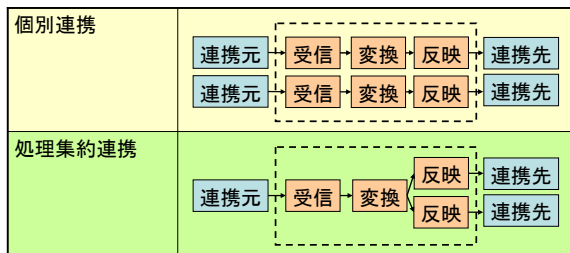


図 3 連携先増加時の連携処理比較

以下、トランザクションあたりのレコード数が 1 と 300 の場合について、連携先が増加した場合の集約効果を机上検討した結果を図 4 に示す。特徴として、連携先数の増加に従って効果が大きくなる傾向が、また 1 トランザクションあたりのレコード数が少ないほど、効果が大きくなる傾向が確認される。例として連携先が 2 連携になった場合の集約効果を示す。個別連携に対して、1 トランザクションあたり 1 レコードの場合は、79%の処理時間となる。一方 1 トランザクションあたり 300 レコードの場合は、96%の処理時間となる。

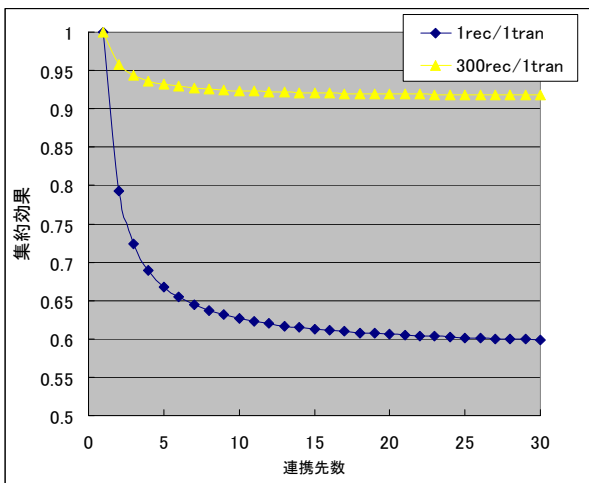


図 4 連携先増加時の集約効果

### 4.2 集約効果が見込まれるデータ連携形態

前節の検討より、トランザクションあたりのレコード数が少なく連携元から複数の連携先に連携する特徴を持つ、データ連携の集約効果が見込まれるデータ連携形態を提示する。

(1) 業務システム間トランザクションデータ共有ある業務システムにて発生したトランザクションデータを複数の別システムにデータ配信する。

(2) マスタデータ配信

統合マスタへの変更を、各業務システムに対してリアルタイムにデータ配信する。

(3) 新システム移行

新システム移行期間に、既存システムとのデータ連携を残し、新システムにもデータ同期する。

### 4.3 今後の課題

(1) 反映処理の性能向上

連携が大規模化した際に、トランザクションあたりのレコード数が大きい場合の集約効果が小さい。トランザクションあたりのレコード数が大きい場合、反映処理に対して性能向上の方式を検討実施する必要がある。

(2) 連携定義の更新

大規模データ連携システムの運用に当たって、連携先や連携元の追加・変更時に、他の連携に從属する連携元システムや連携先システムへの影響を最小限にする方式を検討実施する必要がある。

## 5. まとめ

常駐プロセスによる処理やジャーナルデータ書込み処理といった特徴を持つデータ連携の性能測定を実施、共通化可能な処理集約による連携方式の効果の妥当性を確認した。特にトランザクションあたりのレコード数が少なく、かつ連携元に対して連携先が多い場合の集約効果が大きいことを確認。今後は、大規模データ連携の実現に向け、性能向上の方式検討・評価を実施していく。

## 参考文献

- [1] 楓 仁志, 高山 茂伸, 菅野 幹人: “複数データソース間における効率的なデータ連携処理方式の提案”, 第 73 回情報処理学会全国大会論文集, pp. 529-531, (2011)
- [2] 新井嘉章, 楓仁志, 石井篤, 高山茂伸: ”大規模データ連携処理の実現に向けた実行時設計情報の管理方式”, 第 74 回情報処理学会全国大会論文(2012)