MITSUBISHI ELECTRIC
Changes for the Better

ITソリューションで、エコチェンジ。 eco Changes

Asakusa Framework　適用事例
（業務系バッチ高速化）

2014年11月28日

三菱電機インフォメーションシステムズ株式会社

MDIS

目次

1. 会社紹介
2. 背景・目的
2. 開発概要
3. 成果予測
4. 今後に向けて

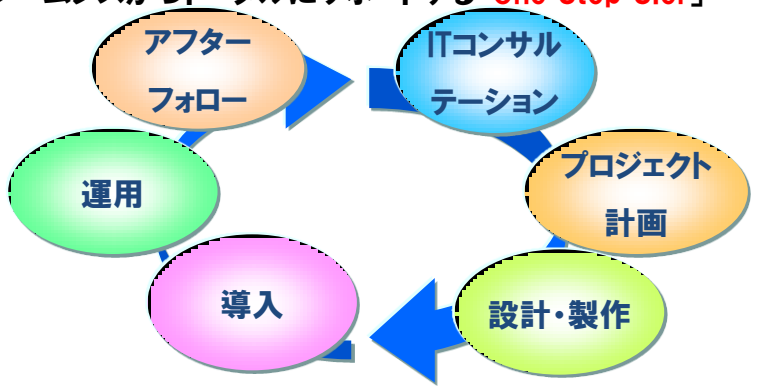
1. 会社紹介

社名　　：三菱電機インフォメーションシステムズ株式会社（略称：MDIS）
本社　　：〒108-0023　東京都港区芝浦 4-13-23　MS芝浦ビル
設立年月日：2001年4月1日
資本金　　：26億円
売上　　：692億円（2012年度）
人員　　：2,184人（2013年3月末）
事業内容　：情報システムの企画設計・開発・製作ならびに販売、ソリューションの提供。
ホームページ：http://www.mdis.co.jp

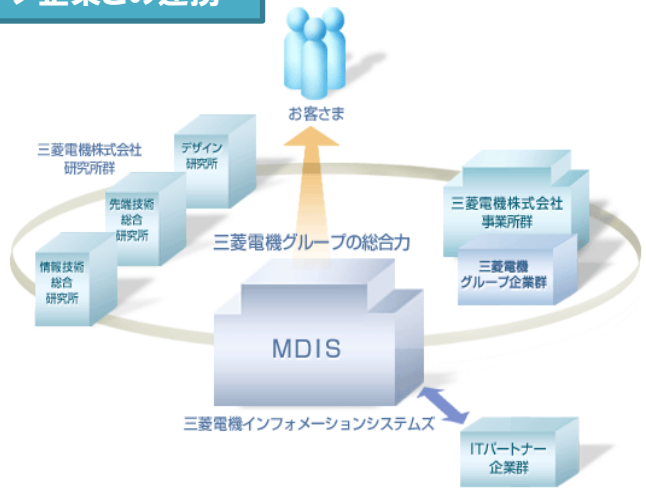
三菱電機インフォメーションシステムズ株式会社（MDIS）は、三菱電機の情報システム事業を引き継ぐ形で、2001年4月に三菱電機から分社化して以降、三菱電機グループのIT事業の中核として、コンサルティングから設計・構築・運用・保守までをワンストップで対応しております。 MDISは、「三菱電機グループの総合力」を最大限に活かし、研究所群やグループ企業群と密接に連携するとともに、三菱電機グループ内にとどまらず、国内外の大手IT企業との強力なパートナーシップにより、お客様に最適なサービスを提供しております。

ワンストップでトータルサポート

コンサルテーションからアフターフォローまでを
シームレスかつトータルにサポートする「One Stop SIer」



グループ企業との連携



2. 背景・目的

（1）背景

従来からオンライン処理は、オープン系へのマイグレーションや再構築が進んでいるが、バックエンドのバッチ処理はあまり手を入れられていないのが現状。

言語の変換、書換え（COBOLやVBからJavaや．Netなど）レベルに留まっている。

業務を支援するシステム化領域の拡大や扱うデータ量、種類の増加に伴う処理量増大の結果

●処理時間や運用、監視にかかる人的コストなどの増加。

●オンライン業務への影響（サービス提供時間の遅れや短縮）

2. 背景・目的

以下のような対応（工夫）、対処療法・・・

●業務に優先度を付けバッチ処理を組立て直し、優先度の高いオンライン業務に関連するバッチ処理を先行させ、他のオンライン業務を遅らせてもらう。

●バッチ処理を日次から週次、月次などへ運用を変更し処理時間を短縮させる事で業務全体への影響を回避。
但し反映タイミングがずれる業務へは影響あり。

●ピンポイントでの改修、処理の分割など

2. 背景・目的

結果・・・

①処理の組立てや運用見直しにより将来に渡って十分対応できている。

②上記で現在は対応できているがさらなる業務量増加や外的要因発生した場合、別の手当が必要と考えている。

③程度の差はあれ、利用部門は我慢しており性能改善してほしい。

②、③への対応

2. 背景・目的

現アーキテクチャ前提で処理を見直す事で現状以上の性能改善の余地はあるが、

a. 業務見直しなどの要因発生した場合、直ぐに別の手当が必要

b. さらなる業務量増加により将来、追加手当が必要

上記対応ではユーザ部門に対し
投資効果の説明が困難

・劇的な改善効果が望める方式
（業務量増加や内容変更にも耐えられる）
・ある程度対応が容易な方式（開発量、費用）

2. 背景・目的

(2)目的(要件)

実行時間が数日レベルのバッチ処理(年次)があり、大量にオンライン(業務)データを更新するため、この期間はリソースを占有し、ユーザへのオンラインサービスは停止

**1日でも処理時間を短縮し
オンラインサービスを
再開させたい**

Hadoop(分散処理による高速化)適用を検討

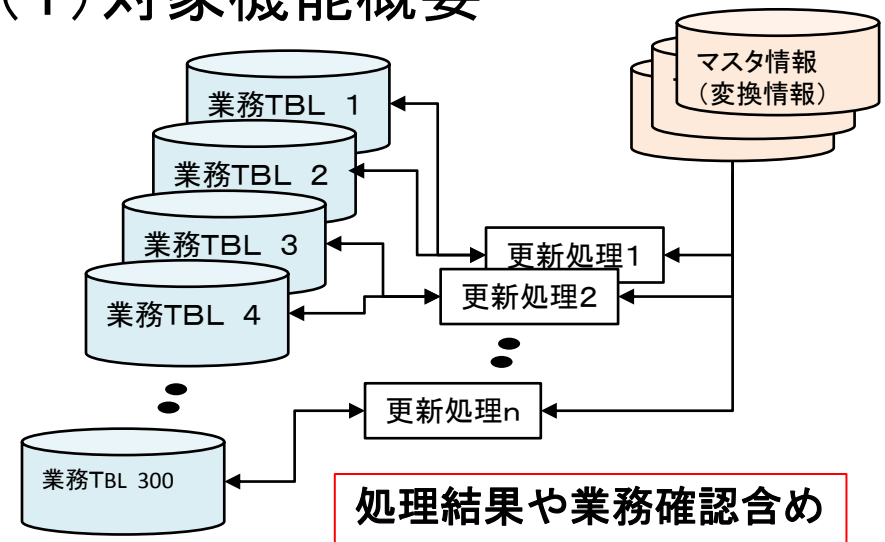
- 大量データ処理(多いほど効果がでる?)
- ボトルネックであるディスクIOに効果

Hadoop適用にあたりAsakusa Frameworkの活用を検討

- 分散上でのデータ配置など意識せず、比較的簡易な命令で開発、難解な?
  MapReduceを意識しない
- データ(の流れ)処理設計に沿った開発を一通り実施すれば他のバッチ処理への適用がイメージし易い、流用可能?
- 機能確認レベルならプラットホーム分散環境不要(人力車)
- (なによりも)OSS

3. 開発概要

（1）対象機能概要



処理結果や業務確認含め
３日程度掛かっている

・処理対象

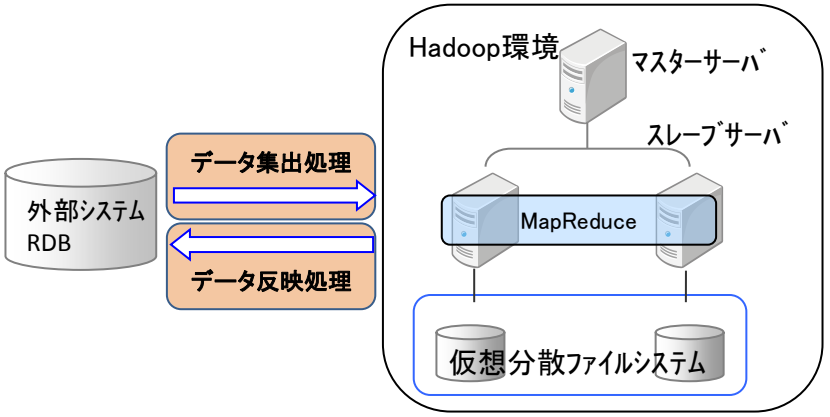
RDB：Oracle10g
テーブル数：約250
全データ容量：約300GB
全データ件数：約8億件
処理データ件数：約7千万件

・処理概要

マスタ情報に基づき、業務TBL 毎に特定項目を更新する。更新処理は業務TBL毎の更新ルールに従い複数存在。

（2）Hadoop化適用処理概要

①前提　HadoopはTEXTベースのファイルシステムを扱い、Oracle（RDB）への直接更新不可のため、前後にOracleDBからHadoop環境への抽出処理と変更後のOracleDBへの反映処理が加わる。



3. 開発概要

（2）Hadoop化適用処理概要

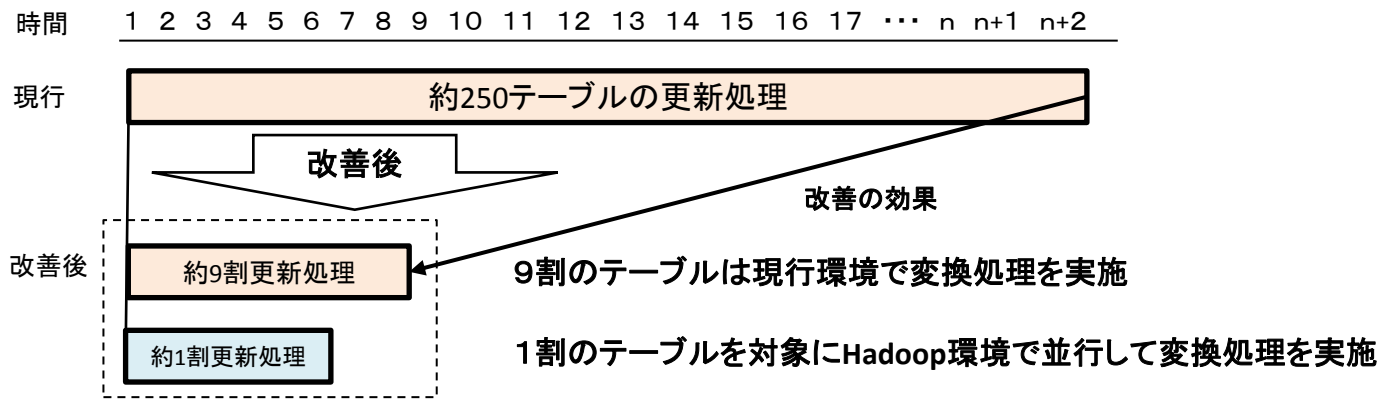
②Hadoop化対象機能選定

更新対象約250テーブルの85%は数秒～10分以内で更新処理が完了しておりHadoop処理前後の処理時間を考慮するとこれらの処理ではHadoop化する事による効果がでにくい、または逆効果と判断。

➤更新時間が一定時間以上掛っているテーブルを対象にHadoop化する事とし、約1割（25テーブル）を対象とした。

※1時間以上更新処理の対象テーブルは約10　最長は4時間半。

③全体イメージ



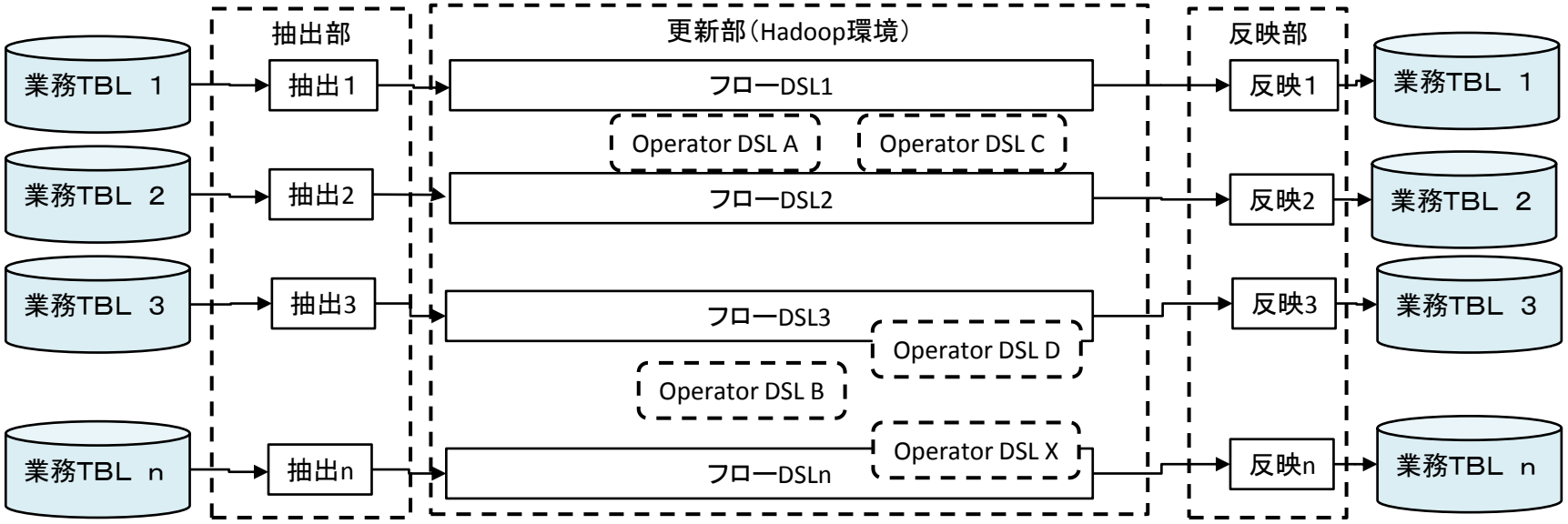
3. 開発概要

（2）Hadoop化適用処理概要

④DSL設計概要

変換対象テーブル毎にJOB（フローDSL）を作成、JOB内の変換処理（OperatorDSL）は変換仕様により共通化。

25テーブル全体で60変換処理（1テーブル2～3変換処理）を23変換処理に共通化。



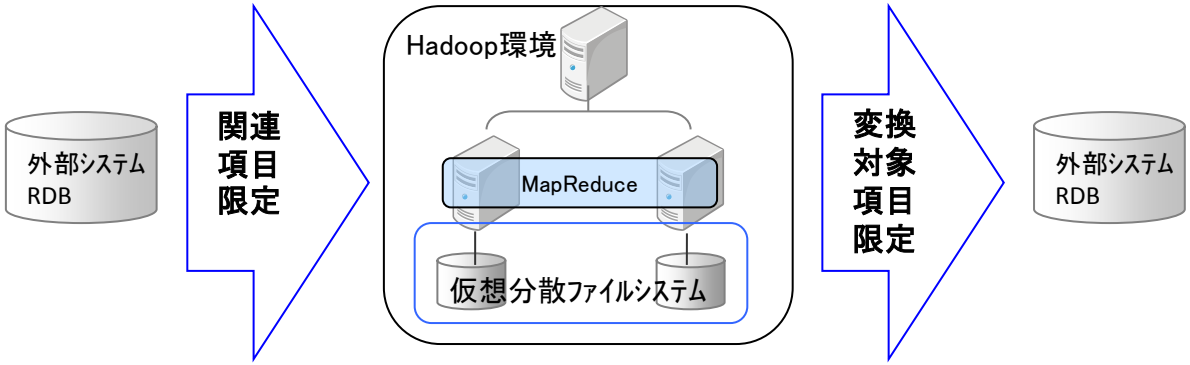
3. 開発概要

本開発にあたり昨年よりノーチラス・テクノロジー社にAsakusa Frameworkを活用するにあたっての手順や設計ポイントをご教授頂き、プロト版の開発を依頼。Map処理が占める割合が高いほど、Hadoop環境での処理性能は格段に向上する事も実感し、同種のバッチをDSLで再設計しHadoop化する事に関しては一通りの成果を上げつつある。

⑤RDB IO設計

今回のような既存RDBの更新処理に適用する場合の重要な設計ポイントはHadoop更新処理前後のマスタ（RDB）とのIO処理。

ポイントは如何にして既存マスタとのIOを少なくするか、必要不可欠な情報に限定しHadoop環境で更新し反映するか。



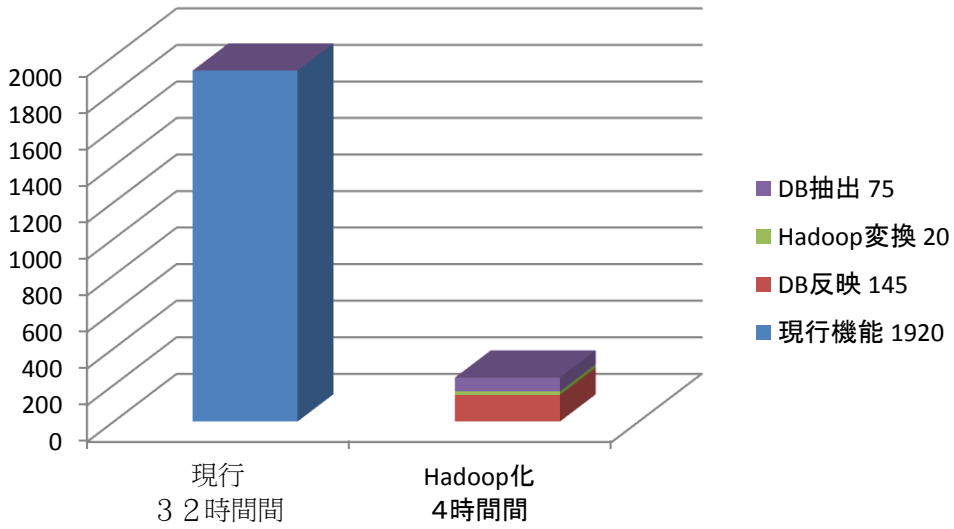
4. 成果予測

Hadoop環境はお客様納入予定の稼働環境、DBサーバは弊社検証環境での測定結果は以下のとおり。

測定対象データ　全データ量：100GB（25テーブル）

全データ件数：430百万件

更新対象件数：36百万件



現行稼働環境：32時間

Hadoop環境：4時間

ノード3台（MapR　M3）

ﾏｽﾀ1台、ｽﾚｰﾌﾞ2台

CPU:XE3-1220v2

3.10GHz

1P 32GB

DISK:500GB

上記より処理全体では、日単位での短縮が可能と判断。

4. 今後に向けて

○基幹系バッチの高速化においては、マスタデータはRDB上に持ち、対象データを一時的にHadoopへコピーして、変換処理を行う形態が主流。

　この為、データ量によりデータ連携処理時間に影響があるため、取扱うデータの内容にあった連携方式の検討が必要。

○本開発は更新系バッチの基本系処理を対象としたが、多様なパターンのバッチ処理への対応が必要、対象データの特性（処理内容、データ間の依存性）や、環境の制約などによる、Hadoop移行パターンの検討が必要。

ご清聴感謝します。

産業・サービス事業本部
グループ事業部
システム第二部　第一課
武石　富士見