

情報統合（EII）の適用

— 情報統合（EII）を有効に活用するために —

FUJITSU ファミリー会 LS 研究委員会「情報統合(EII)の適用 研究分科会」

■ 執筆者 Profile ■



氏 名	団 体 名 (順不同)
郡 和司	田中貴金属工業(株)
山口 大輔	松下電器産業(株)
長野 守史	元・JFE システムズ(株)
磯見 悠	(株)東レシステムセンター
前島 勇	富士通(株)
工藤 久稔	(株)富士通ワイエフシー

■ 論文要旨 ■

企業活動が多くの情報システムで支えられている中で、企業内の IT 情報資産は益々重要性が高まっており、複数システムに分散管理された情報の統合技術は企業価値の向上のために必須である。ここでは、従来から情報統合に使用される物理統合の限界が課題となっており、EII(Enterprise Information Integration)による仮想統合が期待されているものの、効果的に活用されていない状況がある。我々は、EII の「情報のリアルタイム性」「変更に強い」という特徴を企業活動に有効活用するための指針として、1)EII が実際に各企業の業務に役立つかを確認すること（有用性の証明） 2)具体的に EII をどう適用していくかの指針を作ること（適用ガイドライン）を目的として研究を行った。各企業から参加した研究分科会メンバーの具体的な課題をもとに、実機によるプロトタイプを通じて EII の機能・性能両面の検証を行い、期待通りの有用性と定説を覆す性能値を確認した。また、検証で得られたノウハウを適用ガイドラインとしてまとめ、EII 適用の具体的な指針を示した。この成果による情報統合(EII)の有効活用が、ビジネスに貢献することを確信している。

■ 論文目次 ■

1. はじめに.....	3
1. 1 背景.....	3
1. 2 EII に期待すること.....	3
1. 3 EII の課題.....	5
1. 4 研究の目的.....	5
2. 情報統合について.....	5
2. 1 統合方式.....	5
2. 2 EII について.....	6
3. EII の有用性の検証.....	7
3. 1 課題・ニーズ.....	7
3. 2 機能検証.....	8
3. 3 性能検証.....	13
3. 4 検証結果.....	17
4. ガイドライン.....	17
4. 1 ガイドラインの目的.....	17
4. 2 ガイドラインの構成.....	18
4. 3 ガイドラインの使用手順.....	20
4. 4 ガイドラインの検証.....	20
4. 5 ガイドラインの考察.....	20
5. まとめ.....	21

図 1 物理統合.....	6
図 2 共存型.....	6
図 3 EII の仕組み.....	7
図 4 機能検証の環境.....	9
図 5 検証モデル（分散した同種データの統合）.....	9
図 6 メタ定義（分散した同種データの統合）.....	10
図 7 検証モデル（名寄せ表を利用した統合）.....	11
図 8 メタ定義（名寄せ表を利用した統合）.....	11
図 9 検証モデル（ERP パッケージで管理されたマスタの統合）.....	12
図 10 メタ定義（ERP パッケージで管理されたマスタの統合）.....	12
図 11 性能検証の環境.....	13
図 12 性能検証モデル.....	14
図 13 性能測定結果.....	15
図 14 システム構築フローにおける EII 適用ガイドラインの使用範囲.....	17
図 15 ガイドライン・チェックシート.....	18
表 1 ガイドライン・解説書.....	18

1. はじめに

1. 1 背景

1. 1. 1 企業の課題

近年の企業を取り巻くビジネス環境は急激に変化しており、企業内の IT 情報資産はビジネスを行ううえで必要不可欠な要素となっている。しかし、部門や業務ごとに情報システムが構築され、個別最適が追求された結果、情報資産が複数システムに分散管理されている傾向にある。

企業活動において、分散管理されている情報資産の統合は、企業価値向上のために必須である。具体的には、環境の変換に追従できる柔軟な情報システムが求められ、俊敏で正確な企業活動のためには、付加価値の高い情報をより早く正確に収集して活用する必要がある。また、内部統制に対応するためにも、正確かつ信頼できる情報をタイムリーに使いやすい形で活用できる環境の整備が急務である。

1. 1. 2 情報システムの課題

情報システムには、個別最適化されている多くのシステムを統合することにより、全体最適を追求することが求められており、機能面で統合する Service Oriented Architecture (以下 SOA) と、情報面で統合する情報統合が注目されている。

情報に着目した情報システムの課題は次の通り。

- (1) システムの分散により、多くの類似データが分散して存在している。
- (2) 情報連携がバッチ処理中心であるため、新鮮なデータを活用できない。
- (3) 利用者が他システムのデータベース構造を熟知するのは困難なため、必要なデータにアクセスできない。
- (4) 情報統合することで、個人情報や機密情報など利用可能なデータの種類が増えるために、データを安全に活用する手段が必要である。
- (5) マスタの二重持ちによって、ハード面に加えて、マスタ間の矛盾解消、マスタのメンテナンスなど、IT コスト増の要因となる。
- (6) 従来のデータ統合 (Data Warehouse など) では、データ構造が固定化されており、変化への対応力が弱い。

1. 1. 3 課題の解決策

上記の課題を解決するために情報統合は有効な手段である。情報統合には、Extract Transform Load (以下 ETL) による物理統合や Enterprise Application Integration (以下 EAI) を使ったアプリケーション連携が従来から行われているが、これだけでは解決できない課題も多くなっている。これに対して、2006 年度の研究分科会では、企業内情報の統合手段として、Enterprise Information Integration (以下 EII) による仮想統合の効果と可能性が示された。(参考文献[1][3])ここで述べられている EII の「情報のリアルタイム性」「変更に強い」という特徴をうまく活用することで課題が解決できると考える。

1. 2 EII に期待すること

EII の持つ「情報のリアルタイム性」「高い柔軟性」という特徴 (メリット) を情報統合に活用することで、実際のビジネスに対してどのような新しい価値が期待できるのか、3つのケースで考察した。簡単にまとめると、EII に期待する効果は次のとおりである。

(1) 経営層向け情報系システム

企業活動における意思決定では、必要な社内情報を集めて経営/管理層の迅速な判断に使用する。開発納期遵守、販売計画変更、製品市場不良把握、価格変更など

- 情報のリアルタイム性
企業内の基幹システムのデータを直接監視することで、「今」の情報に基づく適切な判断が可能となる。
- 高い柔軟性
刻一刻と変化する意思決定に必要な情報（項目）を柔軟に変更できるため、システム構築を迅速に実施することでスピーディな経営を実現する。
- 新しい価値
企業内の情報を経営層がダイレクトに見えるようにすることで、人・モノ・金・情報といった経営資源を最適に配置し、企業価値を高めることができる。

(2) サプライチェーンマネジメント支援システム

業務ごとに構築されたシステムの情報を横断的に参照することが可能になり、サプライチェーン全体のボトルネック（課題）を「見える化」する。

- 情報のリアルタイム性
各システムで蓄積される情報（生産実績・在庫量・販売実績など）を利用し、それを元にした生産計画・在庫計画・販売計画・補充計画など、より迅速な全体最適化を実現する。
- 高い柔軟性
新しい業務プロセスの追加/削減または変更が行われた際に、影響範囲を最小化し、利用側の参照データの変更も容易にできるシステム構築が可能になる。
- 新しい価値
各部門・企業間での情報を相互に共有することで各業務の個別最適ではなく、ビジネスプロセスの全体最適を実現するための情報活用をナビゲートする。

(3) カスタマーサポート/コールセンター業務支援システム

カスタマーサポートやコールセンターなどのフロント業務に対し、一元管理された情報を提供するシステム構築が可能になる。

- 情報のリアルタイム性
顧客のプロフィールや過去の応対履歴、購入履歴など最新の情報を参照しながら、的確なサービスを提供できる。
- 高い柔軟性
キャンペーンなどの短期サービスを提供する場合など、その都度、利用側の見たい項目が変化するというニーズに対し、迅速に対応できるシステム構築が可能になる。
- 新しい価値
フロント業務担当者は、必要な情報を適切なタイミングで取得し、顧客に合ったサービスを提供できる。このことから、顧客満足度の向上へと繋がりビジネスチャンスの拡大となる。

1. 3 EIIの課題

ここまでで EII の効果と可能性を示してきたが、実際に EII を活用するための課題として次に示す3点がある。

- (1) EII が実際の現場レベルに降りていない。
- (2) EII は検索時間がかかるという印象がある。
- (3) EII に最適な活用ノウハウがない。

1. 4 研究の目的

1.3 で挙げた EII 活用の課題を解決するために、本研究は次の2点を目的とした。

- (1) EII が実際に各企業の業務に役立つかを確認すること（有用性の確認）

具体的な企業の実モデルに対して、EII を適用するプロトタイプシステムを開発し、検証することで、機能面・性能面について実機検証を行い、EII の有用性や課題を確認する。

- (2) 具体的に EII をどう適用していくかの指針を作成すること（適用ガイドライン作成）

EII を現場レベルに落とし込むための、適用ガイドラインとして、実機検証の結果を基に得られたノウハウをまとめる。

2. 情報統合について

情報統合には、仮想統合以外にも様々な統合方式がある。本分科会では、仮想統合の有用性を確認するにあたり、情報統合の方式を仮想統合型、物理統合型、共存型に整理した。

2. 1 統合方式

2. 1. 1 仮想統合型

仮想統合型は、情報源側システムで保持されている実データを、利用側システムの要求に基づいてリアルタイムに収集して仮想的に統合する方式である。

メリットは、情報源側への影響度や情報源側と利用者側への相互依存性が低いことである。また、利用側は最新のデータを参照することが可能である。

デメリットは、動的に統合処理が行われるため一般的に性能に課題があるとされている。詳細は、「2.2 EII について」で説明する。

2. 1. 2 物理統合型

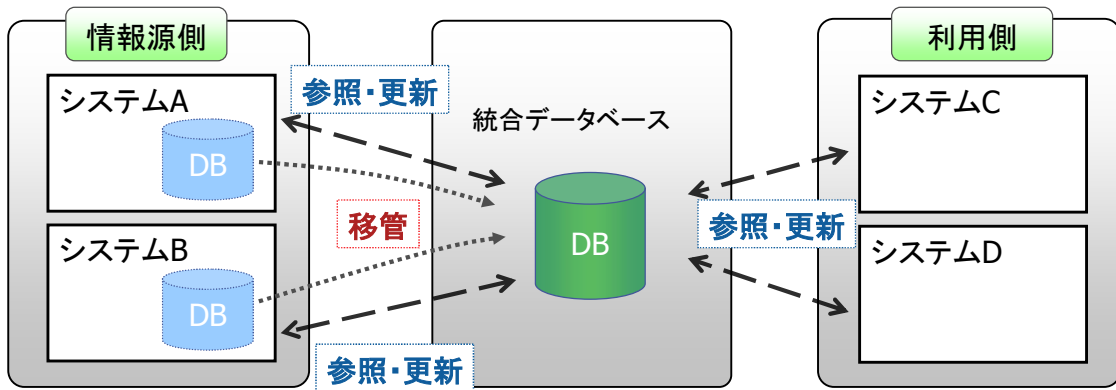
物理統合型は、情報源側で保持しているデータベースを物理的に単一の統合データベースへと移行する方式である。移行後、情報源側及び利用側は、統合データベースをオリジナル情報としてアクセスを行う（図1）。

統合データベースは複数のシステムから使用されるため、そのデータモデルは更新の容易性と、情報の二重管理による矛盾を防止する観点から、正規化を考慮することが求められる。したがって、統合データベースは複数の表に分かれるケースが多い。

メリットは、最新のデータを参照・更新することが可能であることである。

デメリットは、導入時にデータ構造やアクセス先が変わるため既存システムの改修が必要となり、運用時に統合データベースへの負荷が集中する傾向があることである。

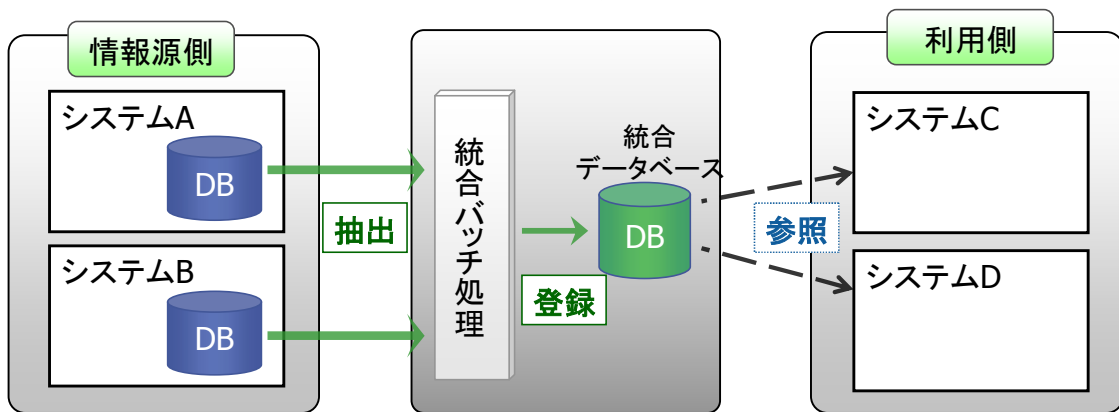
図 1 物理統合



2. 1. 3 共存型

共存型は、情報源側で保持しているデータベース（一次情報）とは別に統合データベース（二次情報）を作成する。統合データベースは、情報源側の一次情報の更新に伴うデータを非同期に受け取り、バッチ的に統合データベースを更新することで、利用側に情報源側の情報を提供する（図2）。

図 2 共存型



2. 2 EIIについて

EII とは、利用側アプリケーションの検索要求に応じて、単一もしくは複数のデータベース（情報源側）に存在する情報から即時に仮想表を作成し、それを検索結果として利用側のアプリケーションに提供することで仮想統合を実現する仕組みである。（図3）

EII の動作はメタ定義で決定され、クレンジング機能により値の表現が統一され、マッピング機能により情報の構造が統合される。

メタ定義として次の3種類の情報を設定する。

- 物理モデル：情報源側から公開される表の性質を設定する。
- 論理モデル：利用側システムから EII を利用する際の仮想表の性質を設定する。
- マッピング定義：物理モデルと論理モデルの関連を定義する。

図3において、利用側のシステムXから仮想表Xに検索を行うと、仮想表Xに必要な情報源に対する検索条件に分解されて、各情報源の検索が行われる。各情報源からの検索結果は仮想表Xに統合され、検索結果として利用側のシステムXに返される。

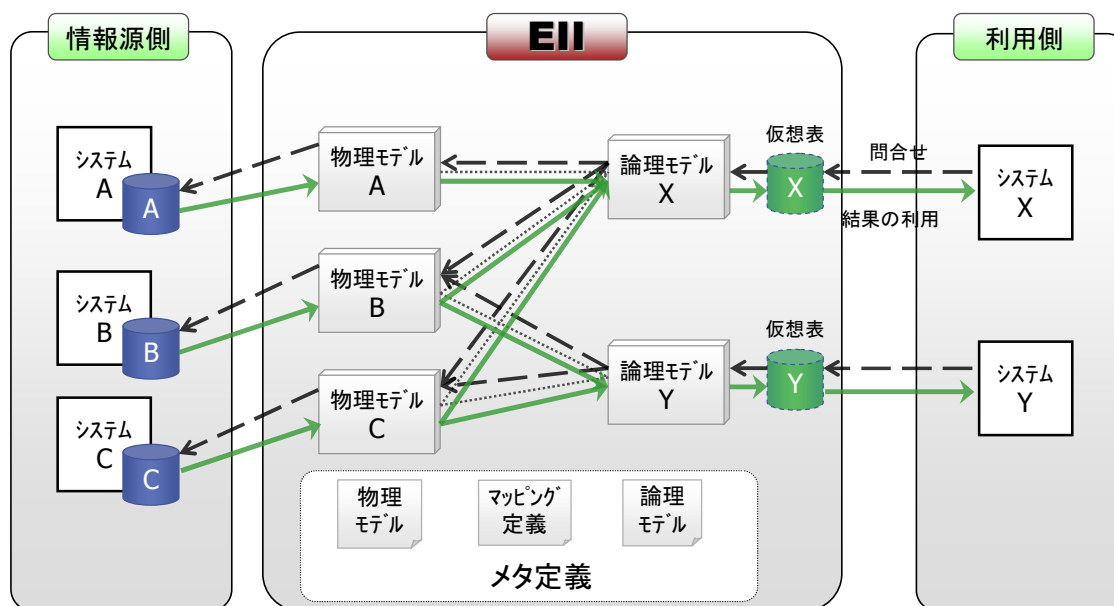
このように、EII では、利用側アプリケーションからの要求に応じて、その都度新鮮な

情報を情報源から取得して情報統合を行うため、情報の鮮度が高い。

更に、次に示す特徴がある。

- メタ定義で仮想表を定義する方式であるため、利用目的に応じた複数の仮想表を定義することが可能であり、多目的利用に適している。
- 仮想表（論理モデル）が情報源側（物理モデル）に影響を与えないため、情報源側のデータモデル及びアプリケーションの変更が不要である。
- 項目の追加や情報源の追加などの変更は、メタ定義を変更することで対応できるため、変化への対応能力が高い。

図 3 EII の仕組み



3. EII の有用性の検証

EII が有用となるのはどのようなモデルであるのか、適用する場合に前提となる条件や、注意しなければならないのはどのような事項であるかを確認するために、EII 適用を想定したモデルを作成し実機で検証した。

検証は、EII の機能が想定どおりのものであるか確認する機能検証、EII で仮想統合にかかる処理時間を確認する性能検証に分け、実施した。

検証結果から、EII 適用が有用なモデル、EII を適用する際に注意する事項を明確にした。

3.1 課題・ニーズ

検証対象のモデルを作成するにあたり、現実的な検証とすべく、本分科会メンバー各社の抱える 6 種のシステム要件の分析から課題及びニーズを抽出した。

(1) 分散したログの統合

SOA 型などの分散したシステムでは、複数サーバにまたがった処理を実行した場合に複数のサーバにログが出力される場合がある。このため、監査などでログを参照する際に、利用者は各サーバからそれぞれログを収集する必要がある。

ログを有効に活用するためには、各サーバに分散したログは単一のサーバに存在しているかのように参照できるようにしたい。

(2) 複数拠点の注文データの統合

複数の拠点で個別に稼働している営業管理システムがあり、工場への生産依頼も地域を固定したシステム連携で行っている。そのため、現状ではシステムを利用して他地域の工場の稼働状況をみて生産依頼を分散させることが難しい。

ここで、各工場の工場注文データを統合して各営業拠点から参照できれば、工場の稼働状況から生産依頼を分散でき、会社全体の生産効率をあげることができる。

(3) ERP パッケージで管理されたマスタの統合

基幹業務を Enterprise Resource Planning (以下 ERP) パッケージで実装し、マスタ管理をメインフレームのアプリケーションで実装しているシステムがある。このシステムでは、メインフレームで更新されたマスタデータを ERP パッケージへ転送し利用している。このため、マスタデータのリアルタイムでの反映ができない、情報の2重持ちによるディスク容量の倍増、といった課題がある。

現行のマスタ管理システムに代わって、複数のマスタを統合して照会、更新できる仕組みを構築できれば課題を解決できる。

(4) 部門別の営業データの統合

営業データを管理するアプリケーションは各部門で個別に運用している。そのため、部門をまたがって営業データを検索する場合には、それぞれのアプリケーションから個別にデータを取得し、利用側システムで集約する必要がある。また、組織変更時に過去データの洗替えをする／しないといった方針が部門ごとに異なるため、過去データを統一的に参照することができない。

部門ごとに異なる方針に影響されることなく部門をまたがった検索ができれば、有効な情報活用ができる。

(5) 複数医療機関の診療情報の統合

各医療機関で独自に管理している診療情報を収集して統合し、医療機関相互で情報を参照できるようにしている。しかし、診療情報の収集、統合作業は手作業で行っており、その中には、医療機関ごとに異なるデータ形式をあわせる作業や、複数医療機関に分散している同一患者の情報は名寄せする作業が含まれている。

このような場合、鮮度の高い情報を参照できないため、リアルタイムで統合できる環境が構築できれば利用者の利便性が高まり、サービス向上につながる。

(6) リサイクル業務データの統合

リサイクル業務の各工程でそれぞれ個別に情報を管理している場合に、取引先からの依頼で顧客対応部門が作業状況を確認しようとしたときには、それぞれの情報を順に検索を行わなければならない。

このとき、各工程の最新の情報が一箇所に統合されていれば、業務全体の管理がしやすくなるとともに、取引先への対応を迅速に行うことができる。

3. 2 機能検証

3. 2. 1 概要

3.1 で挙げた課題・ニーズが EII を利用することで解決できるか、実際にプロトタイプを作成して機能面での検証を行った。

3. 2. 2 検証環境

機能検証では、図4のように、データベースサーバとクライアント PC を同一 LAN 上に配

置する構成で検証環境を構築した。

データベースサーバ2台に4種類の RDBMS をインストールし、各々にデータを分散して格納した。このようにすることで、複数種類の RDBMS の情報を統合する環境を実現した。統合エンジンは各クライアント PC にインストールし、それぞれのクライアント PC で検証が行える環境とした。

尚、本研究で検証に使用した EII ツールは富士通(株)で開発中の Interstage Information Integrator (個別製品版) である。

図 4 機能検証の環境



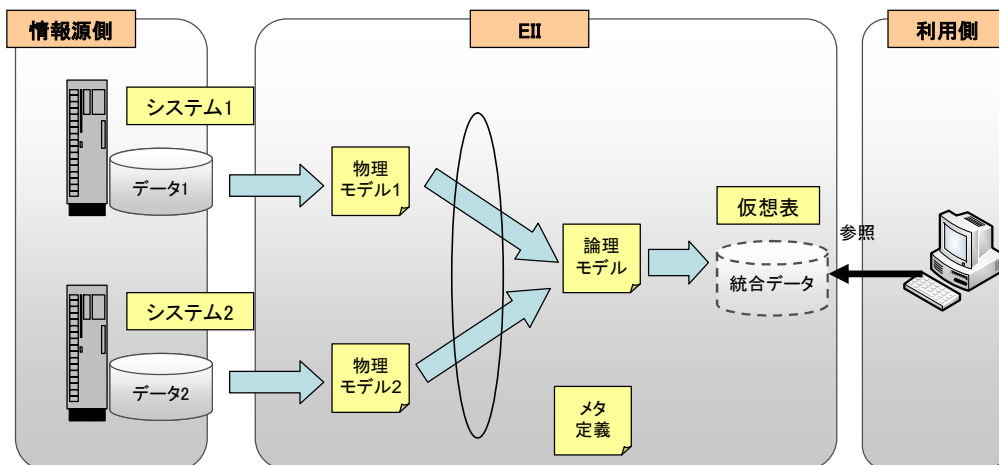
3. 2. 3 検証モデル

3.1 の課題・ニーズから検証ポイントを整理して3種の機能検証モデルを作成した。

(1) 分散した同種データの統合

複数の拠点やサーバに分散している同種のデータを利用側で統一的に参照できるか検証する。

図 5 検証モデル (分散した同種データの統合)

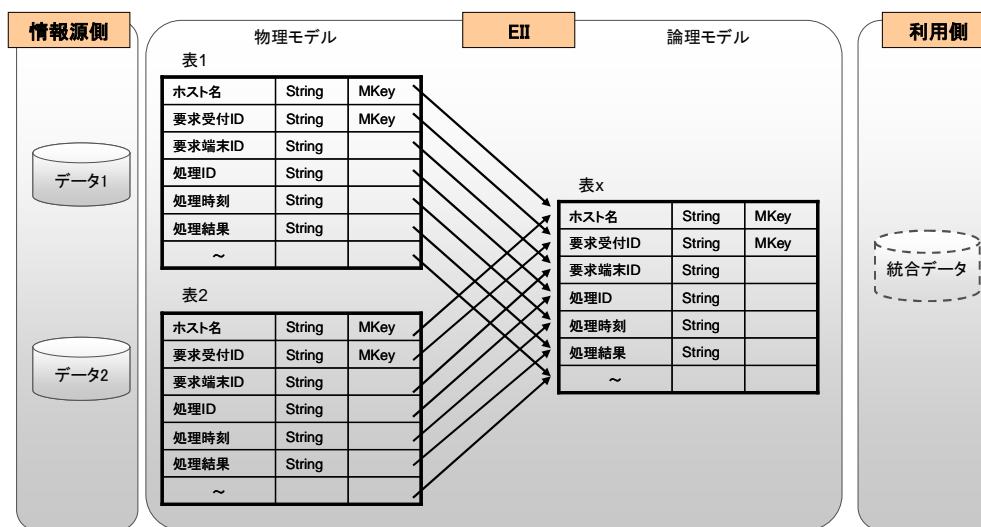


- a. 検証で確認する課題・ニーズ
 - ・ 分散したログの統合
 - ・ 複数拠点の注文データの統合
 - ・ 部門別の営業データの統合
- b. モデルの構成

複数サーバで管理されているデータを EII で仮想統合し、利用側ですべてのサーバのデータが参照できるようにする。(図5)
- c. メタ定義

各サーバに存在する同一フォーマットのテーブルを、EII を通して一つの仮想表として見られるように定義する。(図6)

図6 メタ定義 (分散した同種データの統合)



- d. 検証ポイント

検証にあたり、次の4点について確認した。

 - ・ 複数サーバに分散した情報源のデータを利用者が一括参照できる。
 - ・ 情報源のサーバごとにRDBMSの種類が異なる場合でも統合できる。
 - ・ 利用者が情報を取得し参照する際に、データの並び替え、様々な条件で抽出が行える。
 - ・ 情報源での最新のデータ更新結果が、利用側でリアルタイムに参照できる。

(2) 名寄せ表を利用した統合

分散して管理されている情報に対して名寄せ表を使って統合し、利用側で統一的に参照できるか検証する。

- a. 検証で確認する課題・ニーズ
 - ・ 複数医療機関の診療情報の統合
 - ・ リサイクル業務データの統合
- b. モデルの構成

複数サーバで管理されているデータを EII で仮想統合し、利用側ですべての

サーバのデータが参照できるようにする。このとき、名寄せ表を利用してデータを分類する。(図7)

c. メタ定義

各サーバに存在する同一フォーマットのテーブルを、EIIを通して一つの仮想表として見られるように定義する。このとき、名寄せ表を利用し製品分類情報を付与する。(図8)

図7 検証モデル (名寄せ表を利用した統合)

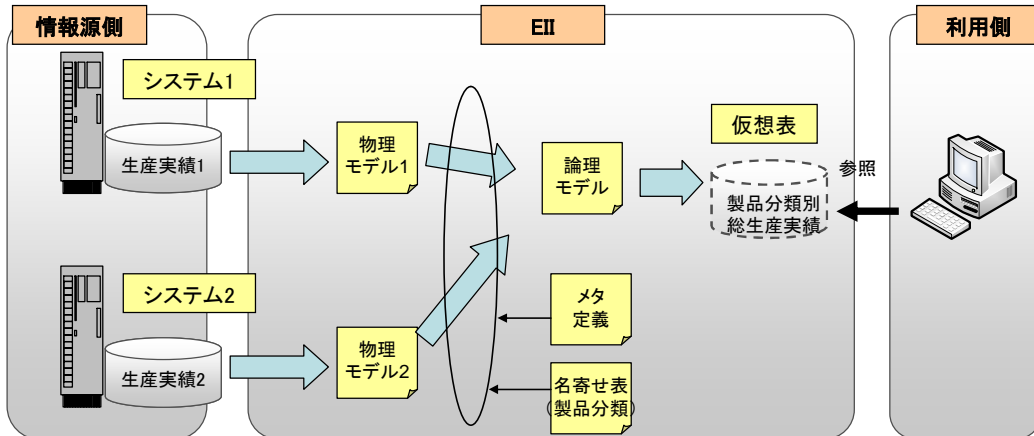
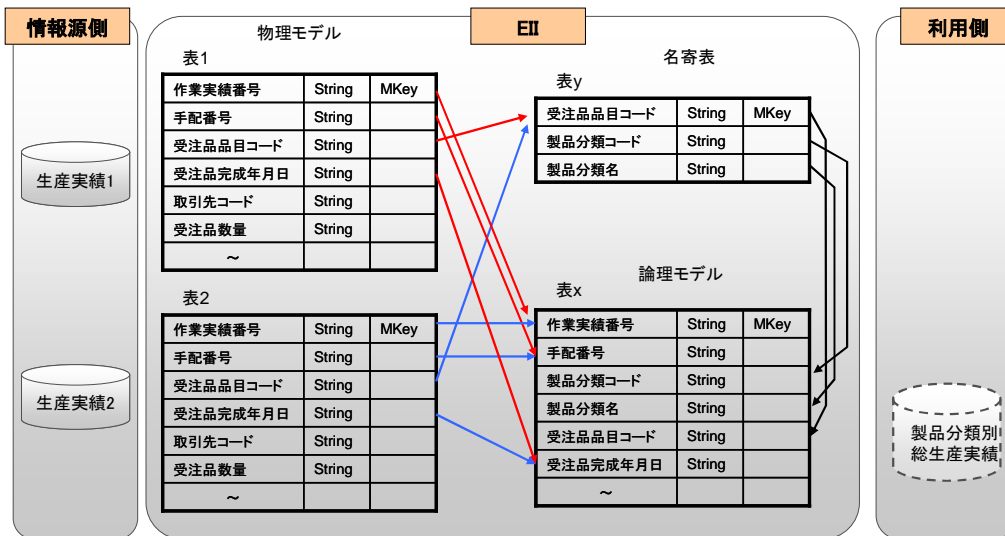


図8 メタ定義 (名寄せ表を利用した統合)



d. 検証ポイント

検証にあたり、次の3点について確認した。

- ・ 複数サーバに分散した情報源のデータを利用者が一括参照できる。
- ・ 名寄せ表を利用して、論理モデルに項目を追加し参照できる。
- ・ 情報源での最新のデータ更新結果が、利用側でリアルタイムに参照できる。

(3) ERP パッケージで管理されるマスタの統合

ERP パッケージで管理されているテーブルを統合できるか検証する。

a. 検証で確認する課題・ニーズ

- ERP パッケージで管理されたマスタの統合
- b. モデルの構成
 - 販売価格、購買価格テーブルを統合し、統一的な価格マスタとして参照できるようにする。(図9)
- c. メタ定義
 - 複数の価格情報テーブルを、EIIを通して一つの仮想表として見られるように定義する。このとき、一度結合した価格データを論理テーブルで分割する。(図10)

図9 検証モデル (ERPパッケージで管理されたマスタの統合)

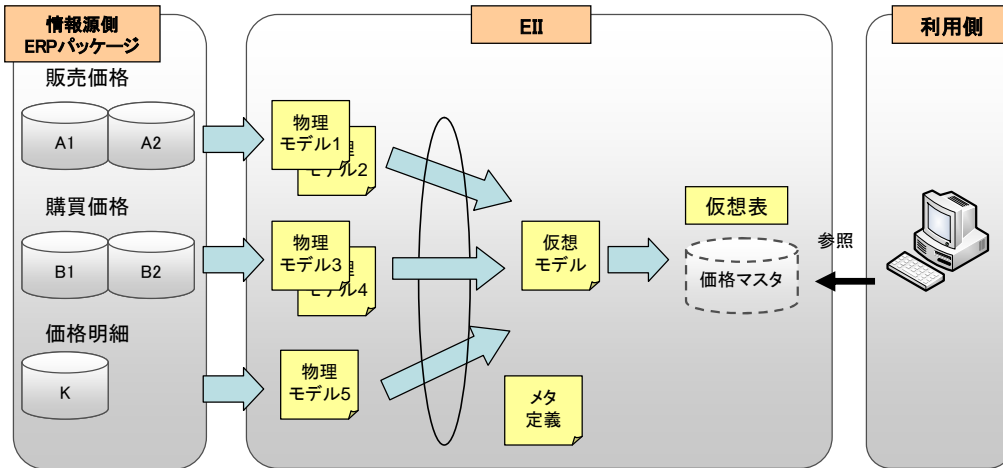
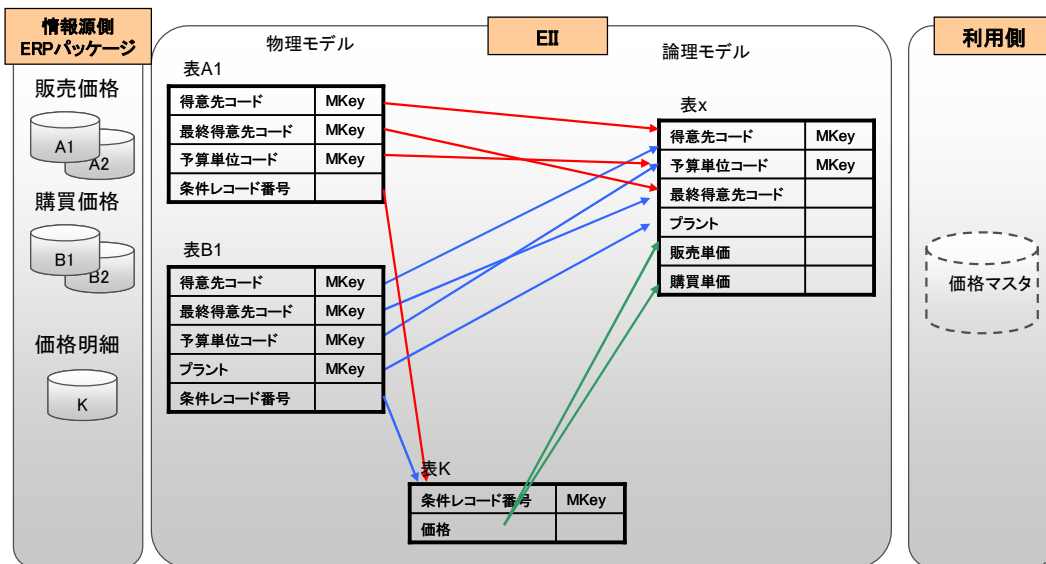


図10 メタ定義 (ERPパッケージで管理されたマスタの統合)



- d. 検証ポイント
 - 検証にあたり、次の3点について確認した。
 - 情報源のマスタデータの二重持ち解消によるディスク容量の削減ができる。
 - 情報源での最新のデータ更新結果が、利用側でリアルタイムに参照できる。
 - ERPパッケージの正規化されているテーブルのデータを統合できる。

3. 2. 4 機能検証結果

これら3種の機能検証の結果から、確認できたこと、できなかったことをまとめる。

(1) 機能検証結果から確認できた内容

機能検証結果から次の4項目を確認することができた。

- a. 情報源のデータが更新された結果をリアルタイムで参照できることを確認した。
- b. 物理的に異なるデータベースの情報をひとつの仮想的なデータベースとして参照できることを確認した。
- c. 仮想統合の環境を構築することで利用するディスク容量を削減できることを確認した。
- d. メタ定義を編集するだけで、仮想表を変更できることを確認した。

(2) 機能検証結果から抽出した課題

機能検証を行った結果、次の2項目を課題として抽出した。

- a. 統合したデータについて並び替えをして抽出できない。
- b. 結合したデータを仮想モデルで分割できない。

上記の課題は今回使用した EII ツールの課題であり EII 自体の課題ではないが、データモデルについては、EII ツールごとに特性がありそうなので注意すべきである。

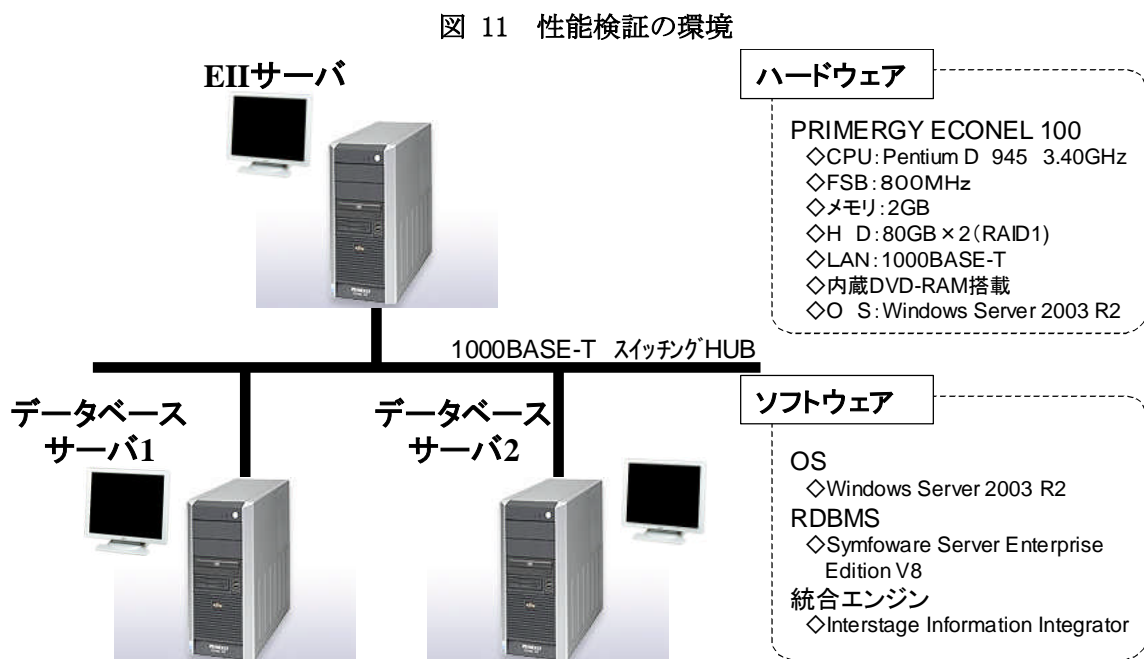
3. 3 性能検証

3. 3. 1 概要

一般的に EII による仮想統合の性能は仮想表への統合処理をリアルタイムで行うため、検索性能に課題があると言われている。そこで、性能の観点から仮想統合の適用範囲を極めるために、物理統合と比較した仮想統合の処理時間（性能）について検証した。

3. 3. 2 検証環境

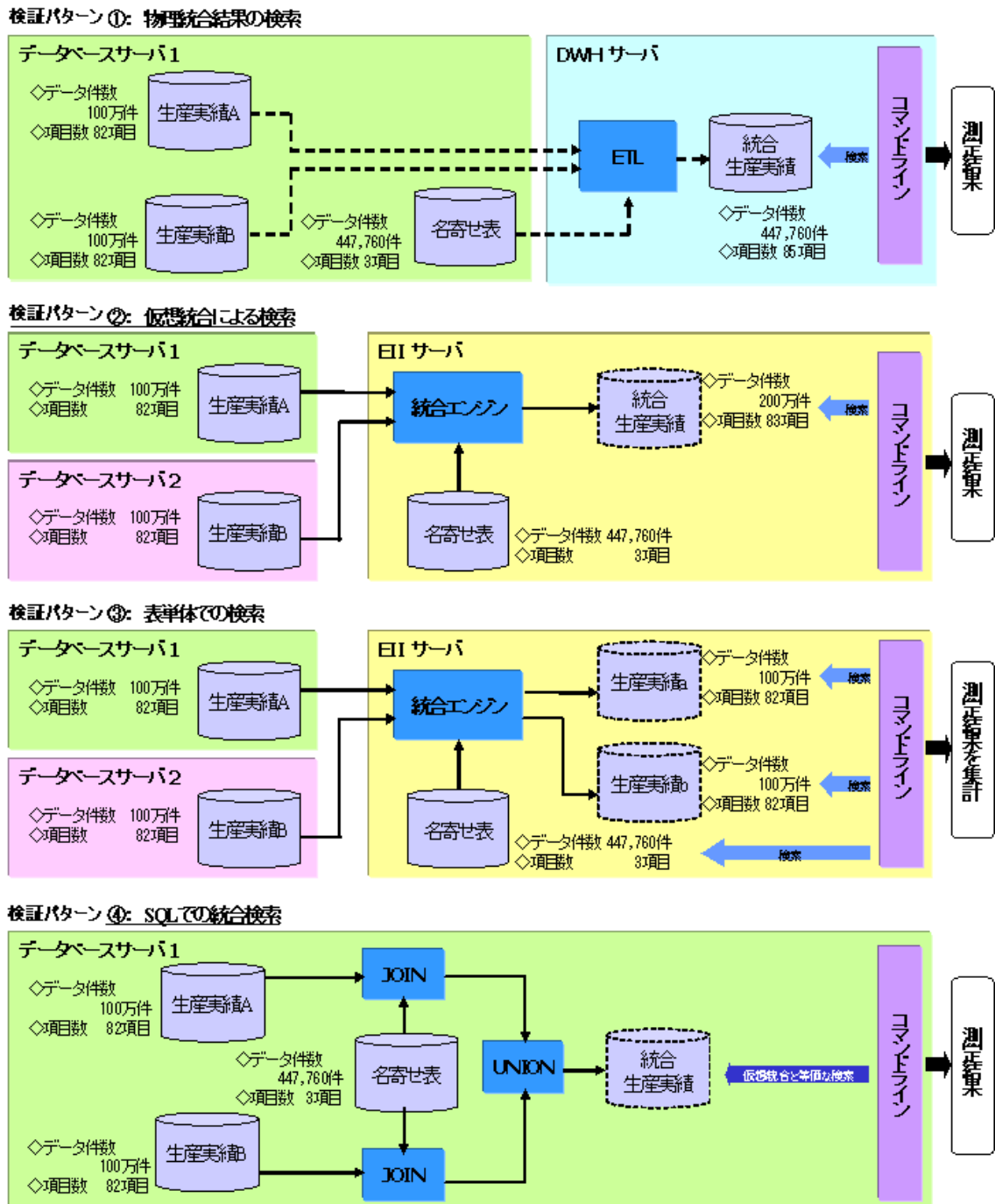
性能検証は図 11 に示すとおり、3 台のサーバに RDBMS をインストールし、内 1 台に EII 統合エンジンをインストールした環境で実施した。3 台のサーバに分散した情報源を EII で統合する検証が行える環境として、実際に近い環境を目指した。



3. 3. 3 検証モデル

検証モデルは、機能検証モデル(2)の基である実際の業務システムを参考にして 82 項目約 250 万件の本格的な検証用データを作成した。これは、情報源であるデータベースサーバ1、データベースサーバ2に異なる体系で存在する生産実績データを、ローカルサーバの名寄せ表を使って統合した結果を、利用側で参照するモデルである。このモデルに対して、図 12 に示す4つのパターンの統合を行い、それぞれの検索(統合処理)時間を計測した。

図 12 性能検証モデル



【検証パターン①】 物理統合結果の検索

生産実績 A,B 及び名寄せ表を、予め ETL で物理統合して構築した統合生産実績表に対する検索時間を計測した。なお、インデックスあり/なしの両ケースを検証した。

【検証パターン②】 仮想統合による検索

生産実績 A,B 及び名寄せ表を、EII により統合生産実績に仮想統合する時の検索時間を計測した。なお、名寄せ表を EII にキャッシュする/しないの両ケースを検証した。

【検証パターン③】 表単体での検索

EII をアプリケーションとみなして、各情報源を単体で検索する時間の総計を計測した。検索条件は検証パターン②で EII が情報源に発行する条件を使用した。

【検証パターン④】 SQL による統合検索

一つのデータベースサーバに生産実績 A,B 及び名寄せ表を格納して、検証パターン②と同等の結果を RDBMS の SQL (関係演算) で導出する際の時間を計測した。

性能検証の目的は、次の3点について明らかにすることである。

- (1) 物理統合 (結果表①、SQL④) と仮想統合②の検索時間差は？
- (2) EII による統合②とアプリケーションで各情報源を個別に検索する場合③の差は？
- (3) EII のキャッシュによる効果は？②

3.3.4 性能測定結果

図 13 性能測定結果

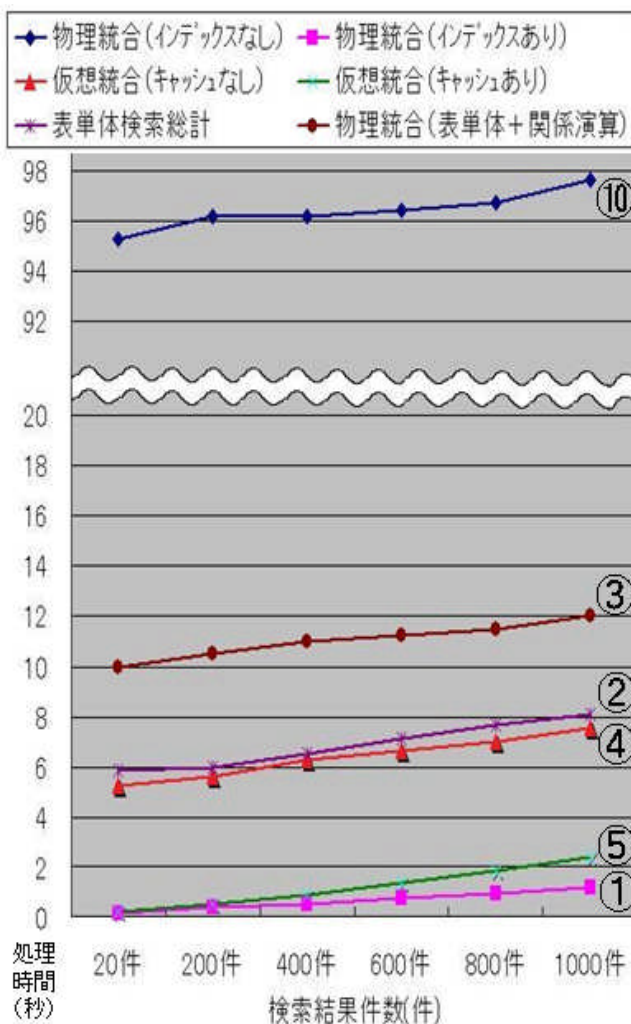


図 13 に示す性能測定結果から次のことがいえる。

- (1) インデックス付の物理統合結果表の検索①が最速だが、インデックスなし⑩は最悪であり、検索時間はどれも結果件数にほぼ比例する。
- (2) 仮想統合④は RDB の関係演算③より高速である。
- (3) 仮想統合④はアプリケーションによる表単体の検索時間総計②とほぼ同じであり、仮想統合処理は微量(高速)である。
- (4) 仮想統合でキャッシュを有効活用することで⑤、インデックス付の物理統合結果表の検索①に匹敵する性能を達成できる。

3. 3. 5 性能検証結果の考察と追加検証

この検証で、仮想統合でも物理統合と同等の性能が出せるという結論を得た。この理由として、次の4点が考えられる。

(1) EII は最適化された検索条件で情報源を検索する

EII は主キー検索の優先や、複合条件の最適化などで、情報源に対する検索が最小になるように最適化された検索条件を作成する。

(2) EII は可能な限り複数の情報源に対して同時に検索依頼を行うため、情報源の検索が多重で実行可能であり、この性質がない物理統合より高速となる場合がある。

(3) EII はメモリ上で処理の中間結果を保持する

EII は中間結果を保持するためのワークテーブルを作成せず、全てメモリ上にデータを保持して処理する。そのため、処理中に行うデータアクセス時間を最小化できる。

(4) CPU 性能の向上

I/O 性能と比べて CPU 性能が飛躍的に向上した結果、物理表の検索時間に対して仮想統合の処理コストが相対的に小さくなっている。

【追加検証 1】(2)を検証するため、別途、すべての情報源を一つのデータベースサーバに格納して仮想統合する実験を行ったところ、表単体検索②より 8.9%遅い結果が出た。仮想統合（分散した情報源）性能④は②より 6.7%早いので、この差(15.6%)が多重処理の効果といえる。

【追加検証 2】(3)に関連して、EII はメモリの容量を超える検索結果となるときに性能が劣化したり処理できないことが考えられる。上記の検証ではこのケースが発生しなかったため、別途、大量の検索結果を求める実験も実施したが、EII のメモリ容量限界に達する前に情報源からの検索がタイムアウトとなるため、限界値の確認はしていない。

【追加検証 3】物理統合結果表①の検索において、キャッシュなしの検索時間が極端に遅い理由について検証した。①の表は 45 万件×85 項目の表であるが、同じ件数で 3 項目の名寄せ表の検索時間と比べると 16 倍となった。一般的に RDB をインデックスなしで検索する際の検索時間は、格納されているデータ量（データ件数×項目数×1 項目の平均データ量）に比例すると言われている。今回の実測性能は、本モデルの理論値（データ量が 28 倍程度）より小さいので妥当な結果であるといえる。

以上の考察から、次の驚くべき結論を得ることができた。

- 仮想統合でも、インデックスを設定した物理統合結果に対する検索と同等の検索性能が出せる。
- 仮想統合は、表単体での検索時間と同等の時間で統合処理できる。

この結論から、本節の冒頭で述べた、「一般的に、EII の仮想統合処理には時間がかかるため検索性能に課題があると言われている」ことについて、必ずしもこの通説が正しいとはいえず、仮想統合でも物理統合に匹敵する十分な実用性能を出せることが確認できた。

この結論は、EII の適用範囲を性能面から拡大できるという意味で重要である。

3. 4 検証結果

機能検証、性能検証の結果から、EII を適用することが有用なモデル及び適用時の留意点についてまとめる。

3. 4. 1 EII 適用が有用なモデル

EII の適用が特に有用なモデルは次の3つである。

- (1) 複数システムに分散管理されているデータを統合するモデル
物理的に異なる環境でそれぞれ管理されているデータを、統一的に参照しなければならない場合に、少ない工数で環境を構築できるため有用である。
- (2) 利用側における統合結果の利用目的が多岐にわたるモデル
利用側で参照したいデータに追加・変更が頻繁にある場合に、メタ定義を変更するだけで対応することができ、迅速な対応が可能である。
- (3) 統合結果のデータにリアルタイム性を求めるモデル
利用側から要求した時点の新鮮なデータに基づいてリアルタイムに統合するため、常に鮮度の高い統合結果が要求される場合に有用である。

3. 4. 2 EII 適用で注意すべき事項

EII の適用における留意点について示す。

- (1) 仮想検索結果のデータが大量になる場合
EII で統合した結果のデータ数が大量になるケースは、処理性能に影響が出る可能性があることが分かった。このようなデータを扱う場合には、ETL など既存技術を併用した環境を構築することや条件を分割した検索で解決を図る必要がある。
- (2) 仮想表による参照だけではなく更新も必要な場合
EII は統合結果の参照が目的なので、EII を介した更新はできない。本来、情報源のデータはそれぞれのシステムで管理されるべきものであり、アプリケーションの関与しないルートで (EII を介して) 更新させることはリスクが大きい。更新は SOA などの別の手段を使って情報源側のアプリケーションの責任で実現すべきである。

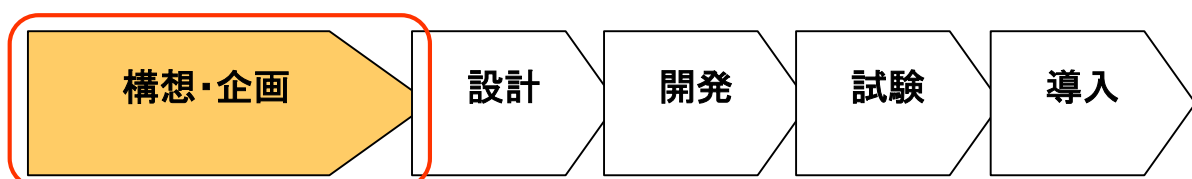
4. ガイドライン

以上で説明した調査結果及び実機検証を基に得た適用ノウハウを、EII の適切な導入を促すためのガイドラインとしてまとめた。

4. 1 ガイドラインの目的

本ガイドラインは、情報統合システムの構想・企画段階 (図 14) でデータ統合の方式を決定する際に使用することを想定し、「情報統合のための方式選定を適切に行う」「適用における注意点とその対処方法を示す」の2つの効果を目的としている。

図 14 システム構築フローにおける EII 適用ガイドラインの使用範囲



A⑧	更新系は SOA などの別方式と組み合わせ使用	<ul style="list-style-type: none"> ・EII は情報源を仮想的に統合するので、EII を介した(仮想表による)情報源の更新には適さない。 ・一般的に、アプリケーションの制御によらない分散トランザクション制御による更新は難しい。 ・情報源を管理しているアプリケーションの制御外で EII が情報源を更新すると、情報源の完全性保障が困難。 	<ul style="list-style-type: none"> ・情報源の更新処理は、情報源を管理しているアプリケーションが全責任を持って実施すべき。 ・SOA で解決できる。→情報源側アプリケーションの機能(サービス)をサービスバスを介して利用することにより更新。
----	-------------------------	---	--

物理統合の注意点		解説	対処
B① B② B⑦ B⑧	性能への配慮が必要	<ul style="list-style-type: none"> ・物理統合は1つの統合 DB を一次情報として、全てのシステムからアクセス(検索・参照・更新など)するため、性能劣化に対する注意が必要となる。 	<ul style="list-style-type: none"> ・統合 DB を分割すると負荷分散できるが、分割が統合目的と相反する場合は分割できない。 ・統合 DB の分割を、情報源側と利用側共に意識する必要がある。
B② B⑧	初期開発コストへの配慮が必要	<ul style="list-style-type: none"> ・情報源システムの DB を廃止して、統合 DB へのデータ移行を行い、全システムが1つの統合 DB にアクセスする改修が必要。初期開発コストへの配慮が必要。 	
B④	運用への配慮が必要	<ul style="list-style-type: none"> ・情報源が停止しても、統合 DB が稼働していれば、利用側は停止前の情報を利用できる。 ・物理統合は1つの統合 DB を全システムからアクセスするため、統合 DB が停止すると、全システムが停止する。 	<ul style="list-style-type: none"> ・統合 DB は無停止運転する。(統合 DB の計画停止には、全システムの計画停止が必要)留意点:統合 DB の二重化は、「データの二重管理をしない」という物理統合の利点と相反する。また、二重化は、可能な限り情報源や利用側が意識する必要が無いように工夫するのが望ましい。
B② B⑤ B⑥	稼働後の改修に対する配慮	<ul style="list-style-type: none"> ・物理統合は統合 DB のデータモデルで全システムが利用する、統合 DB のデータモデル変更を伴うシステム改修は、全システムの改修が必要になる。 	

共存型の注意点		解説	対処
C① C②	複雑な統合処理への配慮	<ul style="list-style-type: none"> ・ETL/EAIの処理が複雑になり次の懸念がある。 (1) 開発・検証コストが高くなる。 (2) 統合DB の二次情報構築に長時間を要する。 (3) 情報源の更新頻度が高いと、運用時の反映処理に長時間を要する。 (4) 統合結果について、対応する情報源が更新された場合の反映範囲やタイミング制御への考慮が必要 	<ul style="list-style-type: none"> ・情報源の更新頻度が小さいケースでは、運用時の処理が少ないため複雑な統合でも性能の懸念は小さい
C① C② C③ C⑩	性能への配慮が必要	<ul style="list-style-type: none"> ・共存型は1つの統合 DB に二次情報を集めるため、統合 DB に参照/更新が集中する傾向があり、性能への配慮が必要である。(更新はミドル経由であるため、物理統合より制御が容易である) 	<ul style="list-style-type: none"> ・二次情報によって統合 DB を分割すると負荷分散できるが、分割が統合目的と相反する場合は分割できない。 ・統合 DB の分割を、利用側及びつなぎ部分(ETL、EAI)が意識する必要がある。
C② C③ C④ C⑩	初期開発コストへの配慮が必要	<ul style="list-style-type: none"> ・既存 DBDB はそのまま、統合 DB に二次情報を集約し、全システムが1つの統合 DB を利用(参照)するため、初期開発コストへの配慮が必要。 	<ul style="list-style-type: none"> ・利用側システム改修: 統合 DB を利用するように改修。 ・統合 DB 設計: 全てのシステムを考慮して DB 設計 ・統合 DB へのつなぎ部分の開発が必要
C④ C⑤	運用への配慮が必要	<ul style="list-style-type: none"> ・情報源の停止時も、統合 DB の稼働により利用側は停止前の情報を利用できるが、情報源が再起動した際に、差分の反映など統合 DB と情報源の整合性を保障する処理が必要となる。(C④) ・1つの統合 DB を全システムから利用するため、統合 DB が停止すると、利用側システムが停止する。(C④) 	<ul style="list-style-type: none"> ・統合 DB は無停止運転する。(統合 DB の計画停止には、利用側システムの計画停止が必要) ・二重化は、可能な限り情報源や利用側が意識する必要が無いように工夫することが望ましい。
C③ C⑦ C⑧	稼働後の改修に対する配慮	<ul style="list-style-type: none"> ・統合 DB のデータモデル変更を伴うシステム改修は、利用側システムとつなぎ部分の改修が必要になる。 	
C⑨	完全な鮮度は実現できない	<ul style="list-style-type: none"> ・情報源(一次情報)の登録と統合 DB(二次情報)への反映に時間差が発生する。 	<ul style="list-style-type: none"> ・EAI で情報源から強制的に統合 DB への反映を行う。 ◇情報源システムの改修が必要。◇利用側システムのレスポンスが悪化◇統合 DB の更新頻度が上がり、性能劣化する。
C⑩	更新系は SOA などの別方式と組み合わせ使用	<ul style="list-style-type: none"> ・統合 DB を利用側から更新すると、一次情報である情報源への反映が必要になるが、実現が困難。 ・双方向に ETL を実装することも考えられるが、情報源側における更新との矛盾解消、情報源側アプリケーションとの分散トランザクション制御が難しい。 ・EAI による情報源への反映処理を実装することが考えられるが、EAI が複雑になり、情報源側における分散トランザクション制御が難しい。 	<ul style="list-style-type: none"> ・情報源の更新処理は、情報源を管理しているアプリケーションが全責任を持って実施すべき。 ・SOA で解決できる。→情報源側アプリケーションの機能(サービス)をサービスバスを介して利用することにより更新。

4. 3 ガイドラインの使用手順

(1) チェックシートの記入

情報統合におけるシステム要件などを考慮した上で、各質問に回答する。

(2) 各統合方式の評価点（○、△、×）の確認

×が一つでも出たら、当該の統合方式は適していない。

△を含んでいる場合は、方式ごとの注意点を確認して方式選定の条件とする。また、注意点とその対処方法についてはそれぞれ解説書を参照する。

(3) 総合評価

×がない方式が複数ある場合は、○が多い方式が適している。ただし、点数の評価時に重点項目の重み付けや△の注意点などを考慮して総合的に判断する必要がある。

4. 4 ガイドラインの検証

このガイドラインをメンバー企業の具体的なモデル事例で検証したところ、6件の全モデルにおいて納得できる統合方式の選定となった。1例について、手順を追って解説する。

(1) チェックシートの記入

各質問に回答した結果、図 15 のようになった。

(2) 各統合方式の評価点（○、△、×）の確認

仮想統合は×がなく、物理統合、共存型は×が一つずつあることを確認する。

- ・ 仮想統合で△がついている質問項目「検索結果のデータ件数が多い（全体を集計した検索処理など）」については「一度の検索で大量の結果となる集計処理を避け、集計処理を分割するなどして1回の検索結果を減らす工夫をすることで適用可能」であることが分かる。
- ・ 物理統合の×が付いている「情報源システムの改修を行える（改修工数、運用面）」という質問の注意点「物理統合には改修が必須」から「本モデルが物理統合には向いていない」という結論になる。
- ・ 共存型で×が付いている「検索元のデータを二重持ちできない」という質問は、大量のデータを倍持たなければいけない場合にハードの増設などコストがかかるが問題ないかということである。共存型の場合は、そもそもデータを二重持ちするという方式のため、二重持ちできないという回答から「本ケースで共存型の適用は向いていない」という結論になる。

(3) 総合評価

×がない方式は仮想統合のみである。上記結果から、集計処理を分割するという条件付きで仮想統合が適しているという結論を導き出すことができる。

4. 5 ガイドラインの考察

4.4 では、本分科会で作成したガイドラインが、実際のプロジェクト構想・企画時に情報統合方式を決める手段として有効であることを示した。

本ガイドラインを作成する過程において、情報統合に関する多数の観点から検討した上で質問項目を最小限に絞り、なおかつ本分科会で得た「各情報統合方式の弱点をどのように克服するか」という知識を盛り込むことで、弱点を克服し目的に合致した適切な情報統合方式を誰にでも選択できるガイドラインを作成することができた。

5. まとめ

近年の企業を取り巻くビジネス環境は急激に変化しており、企業内の IT 情報資産は必要不可欠な要素となっている。しかし、その IT 情報資産は「システムが分散していて、類似システムが多く存在している。」「情報連携がバッチ処理であり、リアルタイムの収集機能や提供機能がないために、新鮮なデータが活用できない。」「従来のデータ統合（DWH など）では、データの種類が固定化されており、変化対応力がない。」などの課題があり、IT 情報資産を有効活用できない事例が多数存在する。

このように、既存技術だけでは利用者の要望に対応できない課題に対して、2006 年度の研究分科会で EII について研究し、その効果と可能性が示された。この研究で示された EII の特徴である「情報のリアルタイム性」「高い柔軟性」を活かすことで、「経営層向け情報系システム」「サプライチェーンマネジメント支援システム」などのビジネスの新しい価値を創出できると考えた。しかし、EII は実際に活用されておらず、最適な活用方法が分からないといった課題が残されていた。

本分科会では、各メンバーの企業内システムで抱えている課題を抽出し、EII 適用モデルを作成し机上検証だけではなく実機での機能検証をすることで、EII が実際の業務で有効に活用できることを確認した。また、性能検証において一般論「仮想表への統合に時間がかかり検索性能に課題がある」とは異なる計測結果を得ることができた。この実機検証から、EII の特徴（メリット）と適用にあたっての注意点を整理することができた。

また、実機検証を基に、「誰にでも目的に合致した適切な情報統合方式を選択できる」ガイドラインを作成することができた。更に、これまでの研究内容を踏まえ、EII を活用することで誰にどのような価値を創出するのか、EII で見出すビジネスの新しい価値を提唱した。

2007 年末から 2008 年にかけて、各ベンダーからもマスタデータ管理に関する製品が次々と発表されている。このような IT 業界の動向から見てとれるように、今後、EII は企業にとって必要な技術になると考える。EII はその柔軟性から、小規模に導入して使用目的によって柔軟に変更していける特徴があり、段階的な導入が可能な技術である。

本分科会の研究により「EII を活用するための準備はできた」と考える。本分科会の成果を活用し、小規模システムから EII を導入することから始め、ビジネスへの貢献に繋がって頂きたい。

参考文献

- [1] 企業内情報の統合による有効活用 ―使ってみましたEII―：2006年度研究成果報告書（第6分冊）、FUJITSUファミリ会 LS研究委員会、2007年5月
- [2] マスタデータ管理ツールの出荷相次ぐ：日経コンピュータ、2007.12.24、20ページ
- [3] E I I による企業内情報の統合：FUJITSUファミリ会2007年度論文 LS研究委員会、2008年2月 <http://jp.fujitsu.com/family/article/2007/summary18.html>