企業間のデータ連携を加速する「iChie」

NTTソフトウェアイノベーションセンタでは，組織間・企業間データ連携による新たな価値創造を促進するため，散在するデータベースを１つのデータベースに仮想的に統合する技術「iChie」（いちえ）の研究開発に取り組んでいます．本稿では，企業間データ連携における課題と，それらを解消するiChieの技術的特長について紹介します．

山本 直人（やまもと なおと）/徳永 大典（とくなが だいすけ）
持田 誠一郎（もちだ せいいちろう）
NTTソフトウェアイノベーションセンタ

企業間データ連携における理想と現実

近年，企業や業界の垣根を超えたデータ連携が注目を集めています．例えば，小売業が持つ販売履歴と，公共交通機関が持つ人流データを掛け合わせることで，商圏分析の精緻化や動線設計の効率化を図ることができます．また，卸売・小売業が持つ販売履歴，メーカが持つ組み立て履歴，サプライヤが持つ製造履歴を掛け合わせることで，製品のトレーサビリティ向上が期待できます．

しかし，このような企業横断的なデータ分析を行う場合，データを１カ所（分析ツールの実行場所）に集める必要があります．また，個人情報や企業秘密を保護するため，データに対して匿名化や秘匿化といった加工処理を施すと，粒度が荒くなり，データとしての価値が低下します．

企業内でも，データベースを部門ごとに導入・運用するケースが多々みられるため，企業間のデータ連携と類似の問題が存在しています．このような企業のいくつかでは，「各データベースから集めたデータをデータレイクにいったん保管し，分析の目的に応じてデータマートとして再構成する」といったアプローチを採っています．構造化・非構造化を問わず，すべてのデータ形式を保存することができるデータレイクを活用すると，各データベースから集めたデータを１カ所で一元的に保管・活用できます．しかし，その一方で個人情報や企業秘密のような機微データを社外のデータレイクに集めることには高い障壁が存在します．そのため，iChieでは，「データベースを企業ごとに分散させたまま，分析アプリケーションに対するエンドポイントのみ一元化することで，仮想的に１つのデータベースとして見せる」というアプローチを採用しています．このアプローチをデータ仮想化（Data Virtualization）と呼びます．

企業間のデータ連携における理想像と現状を表に示します．これらの現状に対して，iChieはデータ仮想化の考え方に基づく特長を備えることで，理想像の実現をめざしています．

ネットワーク特性に基づくデータ転送制御

多くの企業では，ビジネスインテリジェンスツール（BIツール）のようなアプリケーションを用いて，大量のデータを集計・分析し，意思決定に資する情報を獲得します．データを集める際，ビジネスインテリジェンスツールはデータベースに対してSQLクエリを発行します．各データベースはそのクエリに対する応答としてデータを転送します．企業間のデータ連携では，

表　企業間のデータ連携における理想像と現状

	理想像	現状	iChieの特長
1	リモートのデータベースに対して，リアルタイムかつオンデマンドにデータ連携できる	データ転送に時間を要してしまい，リアルタイム・オンデマンド連携ができない	・小さなデータを，大きなデータがある場所に転送することで，データ転送量を削減する ・データベース間のネットワーク品質に基づき，最適なデータ転送経路を選択する
2	データのプライバシや企業秘密を確保したまま，外部からのデータ分析に結果を返すことができる	データ外部に持ち出すため，匿名化，秘匿化などの加工処理を行うと，データ分析価値が低下する	・データオーナーが外部転送を禁止したいデータ項目を指定すると，その項目を外部にデータを転送することなく分析することができる

データベースが地理的，ネットワーク的に分散しているケースがほとんどです．このような状況で，ビジネスインテリジェンスツールが各データベースからデータを集めようとすると，ビジネスインテリジェンスツールとデータベース間のネットワーク品質がボトルネックとなり，データ転送に時間がかかってしまいます．ビジネスインテリジェンスとは，データの組合せを試行錯誤で変えながら，正解を探索する営みです．したがって，データ転送に時間がかかることは，試行回数の減少，すなわち分析品質の低下に直結する問題です．

iChieは，この問題を2つのアプローチで解決します．

1番目は，「小さなデータを，大きなデータがある場所に転送することで，データ転送量を削減する」というアプローチです．従来の方式では，複数のデータベース間でテーブルを結合（JOIN）する際，SQLクエリを発行した場所（1カ所）にデータを集めて結合します．iChieでは，統計情報を利用し，結合対象のテーブルに対する分割クエリにヒットするデータサイズを比較します．そして，ヒットしたデータをSQLクエリを発行した場所に集めるのではなく，ヒットしたデータが小さいデータベースから，ヒットしたデータが大きいデータベースにデータを送って結合処理を行い，結果をSQLクエリを発行した場所に返します．結合されたテーブルは元のテーブルの積集合なので，少なくとも元の大きなテーブルよりデータ量が小さくなります．**図1**は2カ所のデータベース間で結合する例ですが，3カ所以上のデータベースをまたいで結合処理を行う場合は，転送するデータの総量が最小になるように，「どのデータベースのデータを，どのデータベースに，どのような順番で転送するか」という実行計画を立てます．

2番目は，「ネットワーク品質に基づき，最適なデータ転送経路を選択する」というアプローチです．図1のようにデータベース間でデータを転送する場合や，ビジネスインテリジェンスツールにクエリ応答を送信する場合，複数のデータ転送経路が考えられます．iChieは，過去にそれらの経路でデータを転送した際の実効帯域の実績をクエリの実行計画策定にフィードバックすることで，最適なデータ転送経路選択を可能にします（**図2**）．

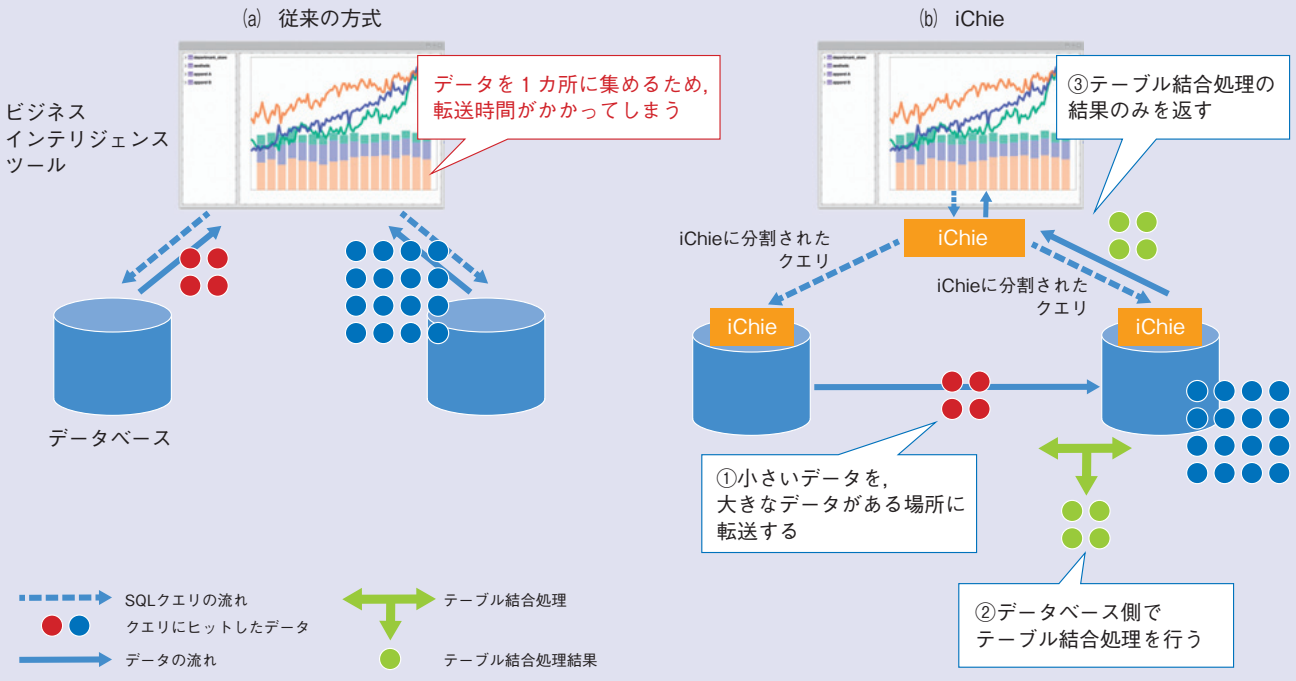


図1　データ転送時間を短縮する仕組み

データオーナによるデータ利用者へのプライバシポリシーの強制適用

iChieはプライバシに配慮した統合分析を支援する機能も備えています．

企業間データ連携で個人情報や企業秘密といった，より安全な管理を求められるデータを取り扱う場合，これらのデータを保護するため，匿名化や秘匿化といった加工処理が必要となります．しかし，匿名加工し過ぎた場合，データの粒度が荒くなり，分析に適さなくなります．

そこで，iChieでは，「プライバシポリシーエンフォースメント」という仕組みを用いて，データを外部に出すことなく，データ分析を行うための機能を提供しています．プライバシポリシーエンフォースメントとは，データオーナ（データベースの管理者）が設定した秘匿化・匿名化ポリシーをデータ利用者に強制適用する仕組みです．

例えば，ショッピングモールが顧客の属性情報を，テナントが購入履歴を持っているケースを想定します．また，ショッピングモールが「会員IDと氏名は外部に出さない」というセキュリティポリシーを持っていたとします．このとき，データオーナであるショッピングモールは，iChieエージェントに対して，「会員IDと氏名は外部に出さない」というプライバシポリシーをあらかじめ設定しておきます（**図3**）．この状況の下，BIツールのユーザが，購入品名ごとの購入日時，年齢，性別を調べたいとします．これを調べるためには，会員IDをキーにして，ショッピングモールとテナントのクエリ応答をJOINしなければなりません．このとき，iChieは，プライバシポリシーが設定されたカラムを持つデパート側のデータベース側でJOIN処理を行うよう，データベースごとのクエリ実行計画を立てます．また，JOINした結果に対して，外部に転送しないカラム，すなわち会員IDと氏名をマスクし，これをBIツールに応答します．このようにすることで，秘匿性が高いデータを外部に出すことなく，データ分析を行うことが可能になります（**図4**）．

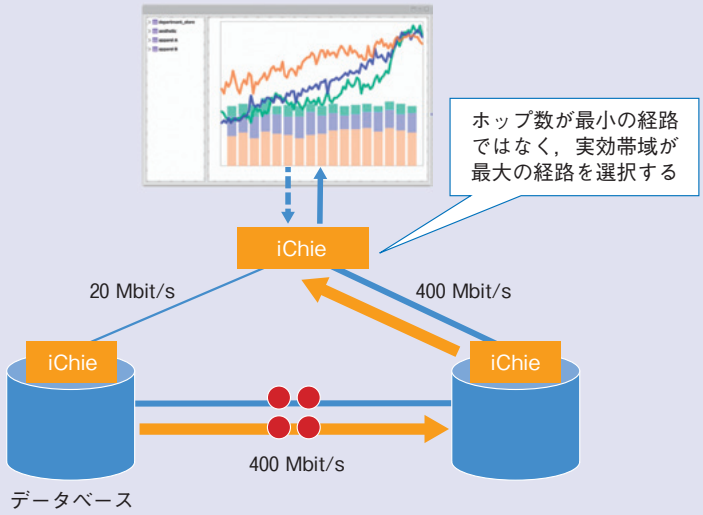


図2　最適なデータ転送経路選択

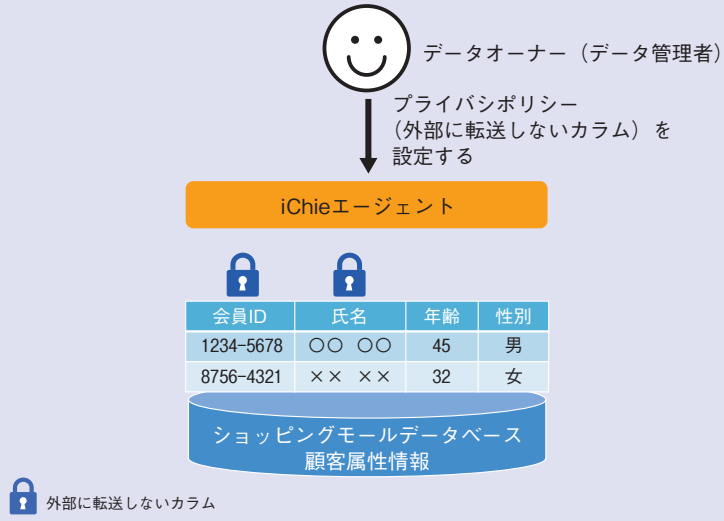


図3　データオーナーによるプライバシポリシーの設定例

今後の展開

2011年，世界経済フォーラムはデ

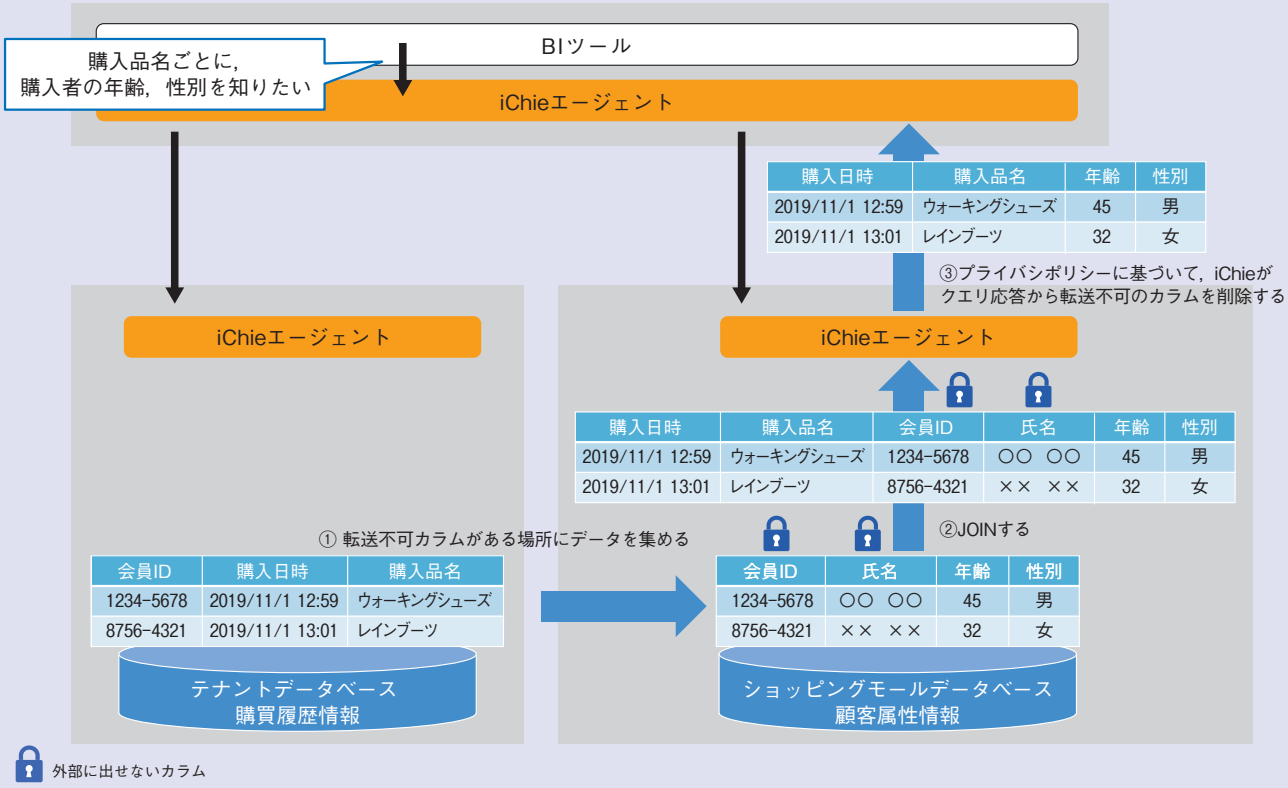
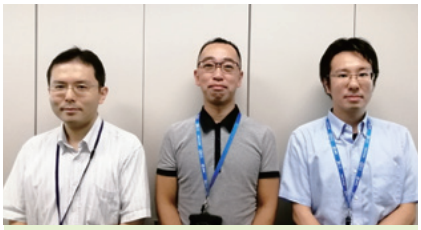


図 4　秘匿性が高いデータを安全に分析

ボス会議で，「21世紀において個人情報は新たな石油（価値ある資源）となるだろう」と報告しました．データは石油のように価値がありますが，それはデータどうしを互いに組み合わせ，適切な分析を施すことによって価値が最大化します．したがって，今後，企業間・産業間のデータ連携に対するニーズはますます高まっていくものと思われます．

企業間データ連携における課題は，本稿で挙げた課題以外にも多く存在します．例えば，企業間のデータ連携では，同一のデータを，同一のカラム名や同一の型で管理しているケースは稀であるため，企業間でテーブルを結合する場合には，名寄せする作業が必要になります．そこで，iChieでは，データのセマンティクスを解釈し，データベース間で共通の表現形式に変換するといった技術を検討しています．また，協調分散機械学習を用いて，企業秘密である実データの共有ではなく，それぞれの企業が自社のデータを用いて生成した学習モデルのみを共有・統合することにより，実データの共有と同じ効用が得られるようにする技術の開発も検討しています．

NTTソフトウェアイノベーションセンタでは，NTTグループのみならず，さまざまなパートナーの皆様と連携し，実社会への適用に向けた取り組みを進めていきます．



（左から）持田　誠一郎/ 徳永　大典 /
山本　直人

NTT研究所では，iChieの研究開発を通じて，組織間・企業間データ連携による新たな価値創造に貢献していきます．

◆問い合わせ先
NTTソフトウェアイノベーションセンタ
第三推進プロジェクト
TEL　0422-59-2822
E-mail　naoto.yamamoto.mf@hco.ntt.co.jp