

## 2 仮想データレイク

企業の垣根と距離の壁を超えた  
多種・大量データの相互利活用が可能に

多くの企業がデータレイクを構築しデータ活用に取り組むようになった。企業の垣根と距離の壁を超えた多種・大量データの相互利活用により新たな価値を生み出すことへの期待も高いが、データ流通や管理に関して課題があり容易ではない。NTTソフトウェアイノベーションセンタ（以下、SIC）はそれら課題を解決する「仮想データレイク」の実現に向けた技術開発を進めている。

企業間のデータ利活用を  
促進する「次世代データハブ」

企業や組織に遍在する多種・大量データを相互利活用し、新たな価値創出につなげることへの期待が高まっている。スマートシティはその代表例と言える。

そこでSICが研究開発を進めているのが、3つの要素技術「仮想データレイク」、「データブローカー」、「データサンドボックス」で構成される「次世代データハブ」だ。本稿では仮想データレイクについて紹介する。

## 仮想データレイクの概要

図1に示すように仮想データレイクには大きく3つの特徴がある。

## 遍在データを仮想的に統合

データそのものを収集するのではなく対象データに関するメタデータのみを収集する点が最も大きな特徴と言える。

「データが遍在した状態で、データ利用者がデータを検索・取得する際には自拠点にデータがあるかのよ

うに体感できる仕組みを実現したいと考えています。そのため、データ更新時にはサブ秒単位のオーダーでメタデータを更新することを目指しています。」（河井氏）

現状、目的毎に構築されるデータレイクがそれぞれデータの複製を持つため、同じデータの複製が多数存在するという状況が起こる。仮想データレイクでは、そうした膨大なデータの複製が不要になり、管理の手間がなくなることもメリットだ。

また、仮想データレイクは繰り返し利用されるデータを効率的に取得できるようキャッシュ機構を持つ。データ転送に伴う処理時間増大を抑え、実用上問題のない時間でのデータ取得を可能にしている。

## データに対するガバナンス

データ提供者が安心・安全にデータ提供できるよう、またデータ利用者が利用可能なデータを簡単に見つけることができるよう、データ提供者のポリシーに基づきアクセス範囲や用途を制限する。

「データレイクには利用履歴や、



NTTソフトウェアイノベーションセンタ  
データ基盤プロジェクト

（左）研究主任 河井 彩公子氏  
（右）研究員 水越 大貴氏

データがどのように利用されているか、派生データの存在や、派生データはどのように作られたのかといったことを把握する機能が求められてきました。そうした要件を全て満たし、企業間データ流通に適用可能な技術はまだ存在しないと言って良いと思います。仮想データレイクでは組織をまたがる利用履歴や派生データの把握も可能にしていく考えです。」（水越氏）

## 遍在データへの統一的なアクセス

組織によりデータの管理方式や形式は異なる。仮想データレイクはデータソース上に存在するデータの一覧をデータカタログとしてデータ利用者に公開する。データカタログは、

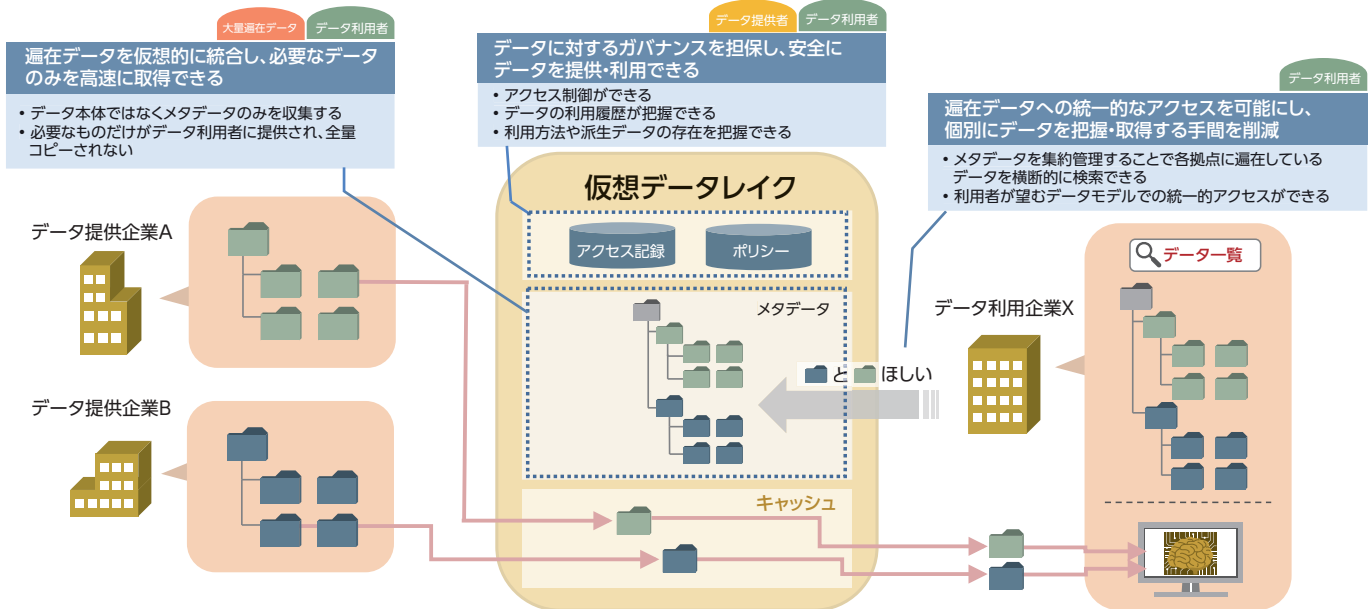


図1 仮想データレイク技術

データ所有者や概要、所在、形式等に関する情報を管理する。データ利用者はデータカタログを介して様々な組織に遍在するデータを横断的に検索・把握し、取得できる。さらに、仮想データレイクではデータ取得のための共通インタフェースを提供し、各データソースの違いを意識せず統一したアクセスを可能にする。

## ユースケース例

### 道路交通の整流化

交通流を最適化する道路交通の整流化に対するニーズが強い。データオーナーが異なり、地理的に分散する交通量データや関連データをリアルタイムに収集・分析し、ドライバー等に迅速に結果をフィードバックするために仮想データレイクが役立つと考えられる。

### リアルタイム電力融通

近年、個人による自家発電も増え

ており、発電量に余裕がある需要家は電力が不足しそうな需要家へ電力融通する需要家間融通が考えられている。需要家の発電量と電力需要に関するデータをリアルタイムに取得・提供し、取引を成立させるために仮想データレイクが役立つと考えられる。

### 製品カーボンフットプリント

製品の生産過程から供給、廃棄、リサイクルに至る温室効果ガスの排出量を可視化する製品カーボンフットプリント（以下、PCF）への注目が高まっている。PCF算出に必要なデータはサプライチェーン全体で多岐にわたり、遍在している。機密性の高いデータが含まれる可能性も高く、データ利用者のポリシーに応じた利用制御が必要なため、仮想データレイクが役立つと考えられる。

## コア技術を段階的に確立中 課題抽出とその解決に注力

技術の一部についてはすでに確立し、プロトタイプを用いた実証実験を開始している。今後の取り組みについて水越氏は次のように述べている。「提供した技術を活用する事業会社との取り組みを通じ、仮想データレイク技術が各種ユースケースにどれだけフィットするかを探ります。同時に課題抽出とその解決も進めています。たとえばキャッシュにデータが読み込まれていない場合、データを要求しても取得に時間がかかるというフィードバックを受け、先読みアルゴリズムをプラグブルに適用するための実装に取り組んでいます。2023年度Q4を目処に現時点で想定している技術を一通り実装した『仮想データレイク v1.0』をプロダクトとして実現したいと考えています。」