

IT インフラレベルのデータ利活用状況の把握を容易化するデータ管理手法の提案 Data Management method to understand IT infrastructure level data utilization

金子 聡[†] 爲重 貴志[†]
Satoshi Kaneko Takashi Tameshige

1. はじめに

近年、データの利活用が進んでいる。データ利活用を行う IT システムは、オンプレミス環境と複数種類のパブリッククラウドを併用するマルチクラウド環境が主流となっている。このように複雑化するマルチクラウド環境において、データの所在や、データ利用状況、運用状況の把握が困難となっている。米国ではデータの適切な削除を怠ったこと理由に 6000 万ドルの罰金が課された事例も起きている[1]。

従来、データ管理ではデータへのアクセスや来歴を管理するが、データ格納先は管理しない。一方、IT 機器管理では、格納先は管理するが、データの中身は管理しない。この管理範囲の分断により、ツール間のデータ突合せや管理者間のやり取りが発生し、利用状況の迅速な把握が困難であった。本稿では、データ利用アプリ層と IT 機器層のデータ管理を連携し、双方のデータ利用を追跡管理するデータ管理手法を提案する。また、社内外のデータ利活用事例の課題を抽出し、提案手法による課題解決の実現性を検討することで、提案手法の有用性を考察する。

2. 従来技術

2.1 Data Management

データ管理の業界団体である DAMA が分類するデータ管理の技術領域[2]のうち、データの利用状況の管理に関する領域は 4 つある。Data Storage and Operation ではデータの格納先を、Data Integration & Interoperability ではデータの移動と集約を、Data Security ではデータのアクセス状況を、Document and Content management ではデータの法令遵守を、それぞれ管理する。これらはそれぞれ対応したツールにより管理されている。データの移動と集約は Data Integration ツールと Data Catalog ツールで管理される。データのアクセス状況は、Data Catalog ツールが、データの利用者や利用条件などのアクセスルールの定義により管理可能である。データの法令遵守に関しては、前記アクセスルールの法令にあわせて設定することで Data Catalog ツールにより管理可能である。これらの管理業務は、Data steward と呼ばれるデータ管理者が執り行い、データ管理の方針に基づいて、具体的なルールを前述のツール上で整備し、データの利用状況を監視することができる。

一方、データの格納先は、ストレージ管理ツールやデータのバックアップツールにより IT 機器の管理者が管理する。そのため、Data Steward は IT 機器管理者と連携して、データ格納先の確認やデータの削除等を行う。

2.2 従来技術の課題

前述した管理者間の管理範囲の分断により、Data Steward はデータの格納先の把握が困難であった。

例えば、Data Integration ツールを使って業務データを分析環境に移動し、分析、加工を行った後に結果を任意の場所に格納するといった一般的なデータ利活用において、業務データは様々な場所に点在することになる。データ管理ツールではこれらのデータ移動自体は把握可能だが、移動されたデータの物理的な格納先はわからない。一方、IT 機器管理ツールではストレージ機器に格納されているデータの内容は管理しないため、データ移動状況はわからない。同様の問題は、IT 機器管理者の運用でも発生する。IT 機器管理者がデータのバックアップのために IT 機器管理ツールを用いてストレージ内のデータを複製する場合、当然データが元とは異なる場所に複製されるが、データ管理ツールは IT 機器管理ツールによるデータ複製を把握していないため、Data Steward はデータの格納先の変化を把握できない。

Data Steward および IT 機器管理者によるデータ格納場所の誤認識は様々な問題を起こしうる。例えば、顧客とのデータの取扱ルールに基づき任意のデータを削除する際に、データ格納先の物理メディアの破壊が義務付けられている場合を想定する。このケースで、データ利活用やデータバックアップによりデータの格納先が不明瞭な場合、データの削除漏れが発生しうる。GDPR などによりデータの格納先の地域の報告が義務付けられるケースでは、虚偽のレポートの結果として巨額の制裁金が課せられる恐れがある。

3. 提案手法

3.1 データ追跡管理方式

データ管理ツールと IT 機器管理ツールで別々に行われるデータの移動や複製等のデータ処理をまとめて管理することで、データの利用状況を統合的に管理するデータ追跡管理方式を提案する。本方式の概要を図 1 に示す。

図 1 の左にある Data Tracking Method が提案方式を示しており、本方式は以下の Step でデータの利用状況を管理する。なお、図内の番号と以下の Step が対応づいている。

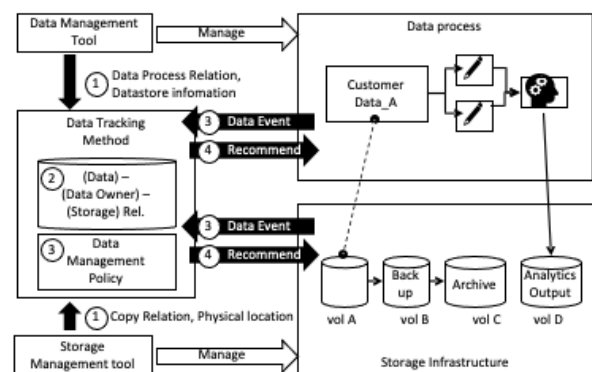


図 1 提案手法の概要

[†] 株式会社 日立製作所, Hitachi Ltd.

- S1. データ管理ツールからデータ (e.g. データベーステーブル) とデータ所有主体 (e.g. DBMS) とデータ格納先のストレージデバイスの関係とデータ移動の情報を, ストレージ管理ツールからストレージデバイスのコピー関係情報と物理ドライブ情報を取得
 S2. 取得した情報を元に各データの格納先を特定
 S3. データ管理ツールのデータ処理イベントと, ストレージ管理ツールのストレージ処理イベントを監視
 S4. イベントがデータ管理ポリシーに影響を与える場合そのポリシーに基づきデータ及びストレージに構成変更提案

上記 S1 と S2 にてデータ格納先情報をデータ処理およびストレージ処理を関連付けて管理することで, データの移動・複製を追跡して格納先を把握できる. また, S3 と S4 では, データの格納先情報を用いて, データ管理ポリシーに影響を与えるデータ処理とストレージ処理を監視し, 適切なデータ運用を提案する. 次節では 2.2 で述べたデータ削除の事例を題材に, S3 と S4 の具体例を説明する.

3.2 データ追跡管理によるデータ運用手法

図 2 にデータ削除のユースケースの例を示す. 従来は, 顧客とのデータ利用規約に基づいて顧客のデータ A (Customer Data A) の削除を求められた際に, IT 機器管理者はストレージに格納されているデータの中身を管理していないので, Data Steward は IT 機器管理者にデータ A の削除を指示できない. そこで, Data Steward はデータ A を現在管理しているデータベースが動作しているサーバを特定し, 対象のサーバに割り当てられているストレージ記憶領域の削除を IT 機器の管理者に指示する. 指示を受けた IT 機器管理者は指定されたサーバの情報からそのサーバに割り当てられているストレージ記憶領域 (volA) を特定し, 削除する. このとき, volA は IT 機器管理者によって IT 環境の管理ポリシーに基づいて, volB へバックアップされ, さらに volC へアーカイブされているが, Data Steward はこれを把握していないので, volB, volC の削除の必要性を認識できない. また, 図に例示するケースではデータ A はデータ管理ツールにおいて 2 種類の加工が行われた後に分析処理が行われ, 分析結果がストレージ記憶領域 D (volD) に格納されている. この分析結果には顧客のデータ A の一部が含まれている可能性があり, 確認する必要がある. しかし, Data Steward はデータの格納先を管理していないため, データ処理後のデータ格納先を把握していない. その結果,

監査時にデータ A の削除を確認されると volB, volC, volD におけるデータ A の削除漏れを指摘される恐れがある.

提案方式を実装したデータ追跡サービスは以下の流れでデータ格納先の確認が行う. 4.1 で示した 4 つのステップに合わせて説明する.

- S1. データ追跡サービスはデータ A が volA に格納された際に, データ A と volA の関係を記憶. さらに, データ A がデータ管理ツールにより前述のデータ処理が行われたタイミングで, データ A が加工された形で volD に格納されたことを記憶. また, ストレージ管理ツールが volA のデータを volB に複製する操作と, volB のデータを volC にアーカイブする操作を監視し, データ A が volB と volC に格納されたことを記憶.
 S2. データ追跡サービスはデータ A が volA, B, C, D に格納されていることを記録する.
 S3. データ追跡サービスは Data Steward によるデータ A の削除操作を検知する
 S4. データ追跡サービスはデータ A が格納されている volA, B, C, D の削除を Data Steward と IT 機器管理者に提案

以上に示した手順により, Data Steward はデータ A が格納されているストレージ記憶領域の確認と削除が可能となり, 監査時の適切かつ迅速な対応が可能となる.

4. おわりに

本研究では, データ管理ツールによるデータ管理と IT 機器管理ツールによるデータ管理の連携により, データ利用アプリからストレージまでのデータ関連を追跡管理するデータ管理手法を提案した. 提案手法は, データとデータを所有する主体とデータ格納先のストレージデバイスの関係と, データ処理によるデータ移動の情報と, ストレージデバイスのコピー関係情報と, ストレージデバイスと物理ドライブの関係と, に基づいて, ユーザのデータ管理ポリシーに適したデータ及び IT 機器の構成変更を提案する. 今後はユーザヒアリングによる課題の妥当性検証とユーザの実環境を対象とした方式具体化を行っていく.

参考文献

- [1] “Morgan Stanley fined \$60m for data center oversight failure,” Data Centre Dynamics Ltd., [オンライン]. Available: <https://www.datacenterdynamics.com/en/news/morgan-stanley-fined-60m-data-center-oversight-failure/>. [アクセス日: March 2020].
 [2] L. Sebastian-Coleman, Navigating the Labyrinth, Technics Publications, 2018.

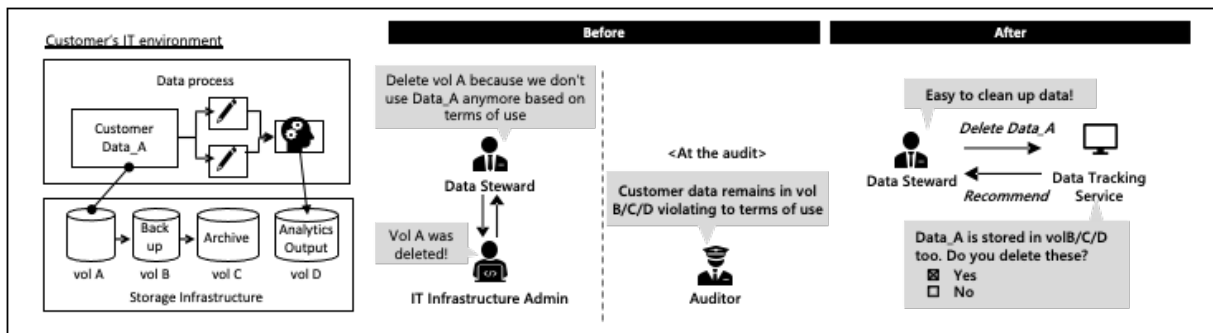


図 2 提案手法を適用したデータ削除ユースケース