

バックアップデータを容易かつ迅速に活用するための データ活用ポータルの開発

水野 和彦[†] 柳沢恵美子[†] 石井 陽介[†] 田中 剛[†]

[†] 日立製作所 研究開発グループ デジタルテクノロジーイノベーションセンタ

〒185-8601 東京都国分寺市東恋ヶ窪 1 丁目 280

E-mail: †{kazuhiko.mizuno.pq,emiko.kobayashi.nv,yohsuke.ishii.bz,tsuyoshi.tanaka.vz}@hitachi.com

あらまし バックアップデータをシステム開発におけるテストデータに活用する等の本来の目的とは異なる用途に活用するニーズが増大している。しかし、データ利用者と管理者が異なることや複数のツール類を利用しなければならないことからデータ準備作業量の増大や誤操作リスクの増大を低減することが課題になっている。そこで、報告者らはデータの要求から提供までの一連の作業を分析し、データ利用環境の準備に必要な情報の一元管理やデータ準備作業の自動化を行うデータ活用ポータルを開発した。このポータルの導入による作業量の削減効果を評価したところ、導入前と比べて約 1/3 に削減可能であることが判明し、データ利活用の容易化・迅速化を実現できる見込みを得た。

キーワード Data governance, data management, backup data, micro service

1 はじめに

業務で発生したデータを他の目的で二次利用する動きが活発化している。例えば、ソフトウェア開発の現場では、システムの品質向上のため、システムテストにおける実際の運用環境の正確な再現のために実際のデータを活用する取り組みがなされてきた [1–3]。最近ではデータベースや仮想マシン (Virtual Machine (VM)) のイメージファイルをバックアップするソフトウェア製品を用いて、バックアップデータをテストデータとして二次利用する運用が提案されている [4]。また、組織内外のデータを分析する分析システムの開発においても実データでテストすることが望ましいと言われている [5]。Jagadish [6] はビッグデータ分析においても分析データの再利用は避けられず、データ収集コスト低減に効果があるとしている。なお、本研究においてバックアップ対象のデータは、データベースだけでなく VM のイメージファイルも含む。

システム開発では、ソフトウェアの開発とその単体テスト、結合テストなど開発とテストを繰り返しながら品質を高めていく。テストの品質を高める方法の一つとして、テストデータとして実際の業務システムのデータを活用する方法がある。テスト環境のデータベースに登録しているデータは、業務システム稼働時のあるスナップショット (断面) データを使用する。既存システムの更新の場合は、更新対象システムのデータを活用する。新規システムを開発する場合、既存システムから抽出されたデータを用いて新規システム用のデータベースを開発することが多い。例えば、既存の決済システムのトランザクションデータを活用して、決済データの傾向分析から商品の発注量を予測するシステムを開発することが挙げられる。このような断面データをテスト専用に入手することはサービスを実際に提供しているシステムから効率よく取得するために、バックアップ

データを活用することが検討されている。バックアップデータは業務システムに故障が発生しデータベースのデータが失われる等の障害が発生したときにデータを復元する手段として利用される。一般にバックアップは、ある決まった時間間隔で取得される。このバックアップを開発環境にコピーし、データのみ復元 (リストア) してシステムテスト等の他の目的に活用する。この復元データの中にはクレジットカード番号のような機微情報が含まれることがあるため、ハッシュ関数を適用するなど解読不能にする処理 (マスキング) を実施してテストデータとして活用する。開発環境を VM 上に構築する場合は、開発ツール類をインストールした VM に、そのテストデータを仮想ディスク等のイメージとして接続してシステム開発者に提供される。

しかし、既存の業務システムのバックアップデータを他の目的に利用する場合、(1) データ管理者とデータ利用者間が異なるため、利用者が期待するデータであるか否かの確認や使用許可を得るまで手続き待ちの発生、(2) 複数の管理ツールの操作が要求される利用環境準備の作業量の増大とイレギュラーな操作であるため誤操作をするリスクが増大する問題が発生すると考えられる。

そこで我々は、これらの問題 (1)(2) を解決するデータ活用ポータルを開発する。データ活用ポータルでは、問題 (1) については、管理者と利用者間の人手が介入する手続きを簡略化するために、データ利用者が自ら希望するデータを選択できるようにするセルフサービスのインタフェース提供と、提供処理の自動化を実現することで解決する。そして、問題 (2) については、バックアップデータに関する情報を一元管理することで、利用するデータの特定やプロファイル情報の入手を容易にし、複数の管理ソフトウェアの設定内容のテンプレート提供や、入力項目数校込み、設定順序の単純化により、データのマスキングや利用準備にかかる管理者の作業量および作業ミスを軽減する。

そして、データ活用ポータルの導入によりバックアップデー

タをシステム開発環境として利用可能になるまでの作業量が削減できることを示す。さらに、この開発環境の再利用や削除作業の観点でも作業量が削減できることも評価する。

本研究の貢献は以下の通りである。はじめに、実際のシステム開発の現場調査に基づいたバックアップデータをシステム開発のテストデータとして活用するユースケースの立案と課題を抽出する(第2章)。次に、これらの課題を解決する方法の立案と、それを実現するデータ活用ポータル提案をする(第3章)。最後に、机上にてバックアップデータを二次利用可能にするまでの作業量を評価する(第4章)。

2 バックアップデータを活用したシステム開発環境の課題

2.1 システム概要

業務システム開発環境に業務システムのデータ等のバックアップデータを活用する事例について図1を用いて説明し、その運用上の課題について述べる。

実際に業務を提供している本番環境のデータベースのバックアップおよび、本番環境のデータを利用する開発環境のデータベースのバックアップデータはともに Actifio [7], Veritas [8], Veeam [9] のようなソフトウェア製品で管理されている。以下、本番環境のバックアップファイルのデータを開発環境で活用するケースをデータ(データベースおよびVM)のライフサイクル順で説明する。まず、本番環境で取得されたバックアップファイルを開発環境にコピーし、開発環境でのテストデータとして活用される。この環境では、データへのアクセス権限のグループとして利用者と管理者を想定する。利用者は、管理者が準備したバックアップファイルを参照可能であり、バックアップファイルを開発環境のVMに接続して利用する。管理者は、利用者の要求に合ったバックアップファイルの検索と、バックアップから復元したデータベースをデータマスキングサーバに接続し、利用者に公開できるようにレコード内の機微情報のマスキングをする。このマスキング済みデータが格納されたデータベースをVMから切断し、システム開発用のVMに接続しなおし、このVMを利用者に提供する。そして、利用者が過去に管理者から提供されたデータやVMをプールして再利用する。最終的には、管理者が開発環境で不要になったVMやバックアップファイルを定期的に削除する。

2.2 運用上の課題

バックアップデータを開発環境で活用するシステムの運用で発生する問題を列挙する。そして、各問題に対し、実際の開発環境を運用している技術者の意見を踏まえ解決すべき課題についても言及する。

【課題1】データが利用可能になるまでの待ち時間の増大 バックアップファイルは本来、データの保護目的であり、別の用途で活用する利用者はバックアップファイルを管理する者とは異なることが多くかつ、非定期的なオペレーションになるため、利用者がデータを利用する際には、バックアップを管理する管

理者の許諾を打ち合わせや申請書のやり取りを通して得なければならない。そのため、利用者が利用したいときにすぐに利用できるとは限らず、データ利用までに時間がかかるという問題がある。この問題については、利用者と管理者が共通のデータ管理基盤上で利用申請とその許可をすることで迅速な処理が可能になると考えられる。そこで、このような共通データ管理基盤を構築し、管理者と利用者に適切なインタフェースを提供することが課題である。

【課題2】利用したいデータの特定作業時間の増大 管理者にとっては、本番環境のデータベースデータを対象としたバックアップまでが定常的な業務であり、開発環境での本番データの利用に対応する作業は定常業務外の作業となる。利用者からの要求に対してデータを準備するには、要求にあったデータを検索し、利用できる状態にするといった作業が必要であり、踏まなければならない手順も多いため作業に時間を要する。また要求にあったデータをすぐに見つけられるとは限らず、利用者への確認作業も発生し、データの準備に時間がかかるという問題がある。VMイメージファイル等の再利用や削除では利用用途、VMを使用した開発プロジェクト情報、取得・作成日、保存期間、利用状況、データのマスキング適用の有無等の様々な条件で検索を行う必要があるため同様に特定作業に時間がかかる問題がある。この問題に対応するために、バックアップファイルやVMイメージファイルに関わる情報の一元管理により、必要なデータを容易に検索、確認が可能にし、これらのファイルのプロフィールや運用設定などの情報をタグ付けすることで、詳細情報の共有を可能にする。これにより、利用者側で希望するデータやVMの要件を入力して提供すべきバックアップファイルやVMを特定できるようにすることで管理者の作業数を削減することが可能と考えられる。つまり、情報の一元管理とタグ機能の実現が検索高速化のための課題である。さらに、利用者が自ら利用するデータを選択できるようにするセルフサービスのインターフェース提供と、処理の自動化を実現することが課題である。このとき、検索対象の絞り込みと情報保護の観点で、利用者には参画している開発プロジェクトに関わる情報のみ見えるようにフィルタリングすることも求められる。

【課題3】管理操作の複雑化 開発環境の管理操作においては、VMの生成や削除などの制御を行うVM管理ソフトウェアおよび、VMに対するデータベースファイルの接続/切り離しやバックアップ取得などを行うバックアップソフトウェアを管理者は何度も両者の間を行き来した操作をするため、操作手順が複雑化し誤操作や実行手順ミスを引き起こす可能性がある。この問題については、複数の管理ソフトウェアの情報を一元管理により確認可能にし、複数ツールをまたいだ操作を自動化し、開発者の作業量を軽減できると考えられる。つまり、管理ソフトウェアの情報の一元管理と操作の自動化が課題である。

【課題4】開発環境保全のための作業量の増大 開発環境で活用しているVMおよびバックアップファイルのコピーの再利用をする際には、利用中のファイルを保全するためのバックアップファイルを取得する。しかし、別の利用者が作業中であるため、開発システムの静止点と判定できるタイミングを容易に決

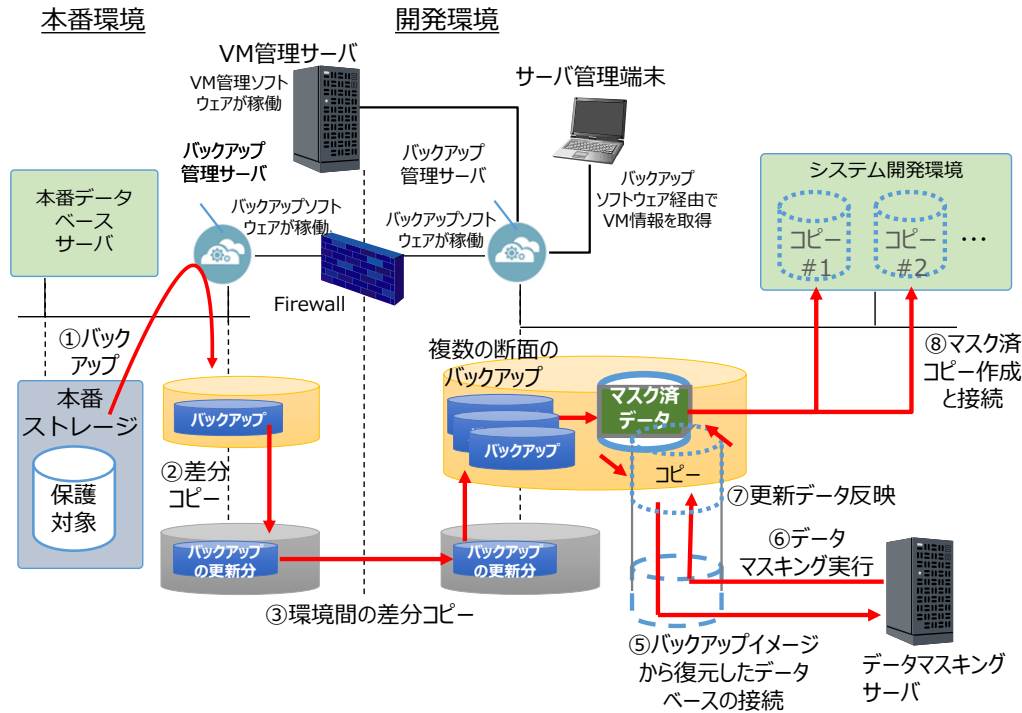


図 1 既存のシステム開発環境

定できない。再利用データを定期的に削除する運用作業をする場合、削除対象となったバックアップファイルのコピーや VM イメージファイルを任意のタイミングで削除すると実運用に影響を及ぼす可能性があるため、これらのファイルの利用者に都度問い合わせで削除可否を確認する必要がある。この問題に対しては、利用者への操作対象ファイルについて利用者への個別に都度問い合わせを無くし、操作を一括処理することで作業量の増大を削減することが可能と考えられる。つまり、この問題に関しても情報の一元化と操作の自動化が課題である。

3 情報の一元管理と作業自動化のためのデータ活用ポータルの提案

データ利用環境の準備に必要な情報の一元管理、作業の自動化を実現するためのデータ活用ポータルを導入したシステム開発環境の概要を図 2 に示す。

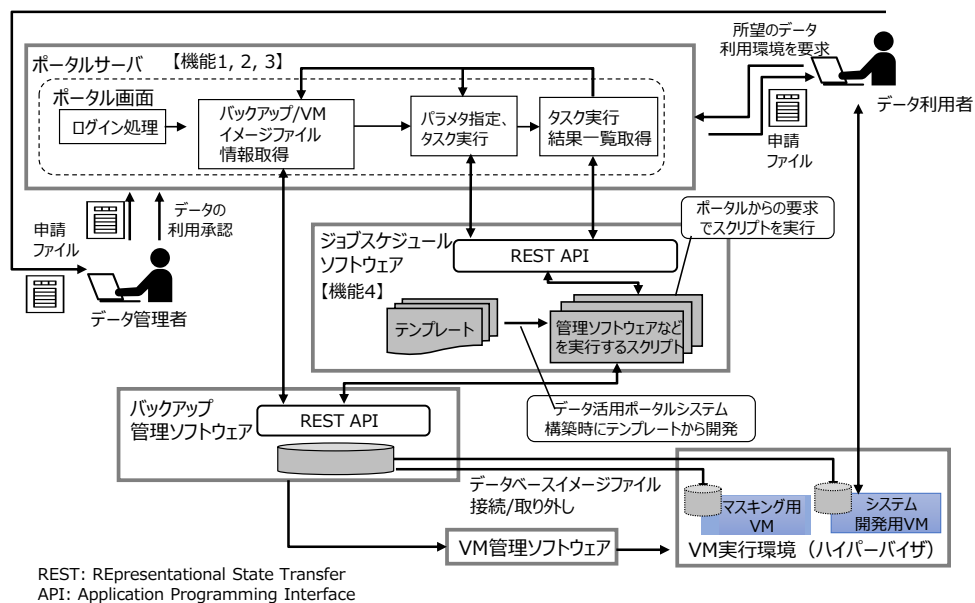
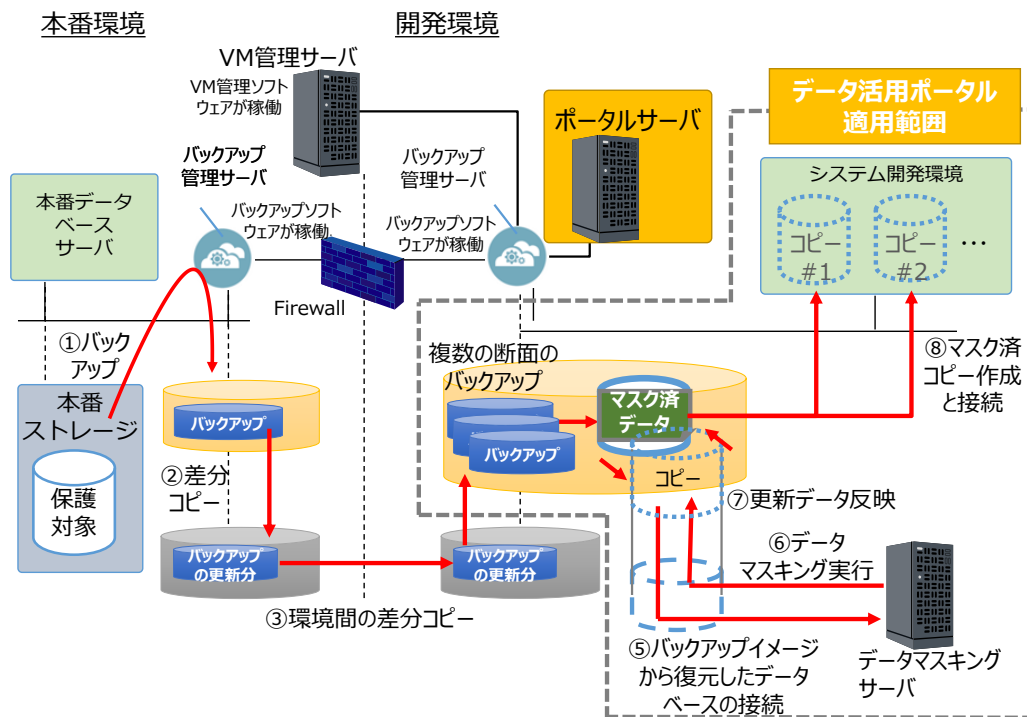
データ活用ポータルのアーキテクチャを図 3 に示す。基本的にはジョブスケジュールソフトウェア (例えば, Ansible AWX [10]) が、ユーザインタフェースでの入力を契機に、データ提供までのシナリオを自動実行する。このポータルの導入で、データの検索にかかる時間、準備にかかる作業工数を削減と開発者のデータアクセスにおける誤操作リスクの低減を実現する。以下に、提供する機能と課題解決へのアプローチについて説明する。

【機能 1】ロール別アクセス制御 ポータルでは利用者からの要求時に、利用者が利用可能なバックアップデータについてのみ、データの情報を出力する。ここで利用可能なデータとは、管理者が許可している種類やマスキング等の必要な加工が施された後のデータである。データ情報の出力については、ポータル

ルでは利用者と管理者とを区別し、アクセスできるデータ情報を制御する。利用者と管理者といった権限 (ロール) 定義をし、ロールに従った出力の制御を行う。ポータルでは、ポータル画面をロール別に用意し、ロールの異なる操作を完全に分離する。

【機能 2】データに対する情報の付加と管理 ポータルではデータを利用者に提供するまでの操作をログファイルに出力するとともに、データのメタ情報として、マスキング実行時にマスキングスクリプト名、およびマウント時にマウント先ホスト名を付加して管理する。また、バックアップイメージファイルとして取得したデータにはデータの用途 (システム名やプロジェクト名など) をバックアップ取得時に、メタ情報として付加する。利用者によるデータ選択時には、バックアップに関する情報と付加したメタ情報 (用途、マスキングを実行するプログラム名) で検索を可能にする情報をポータル画面に出力する。また、管理者が利用依頼条件に一致するデータの検索する時も同様に、メタ情報を含むバックアップイメージファイルのプロファイル情報を検索対象として再利用し、これらの条件指定による検索を可能にする。

【機能 3】利用申請のサポート 管理者側での操作の簡単化のために、利用依頼用のポータルから利用者が自ら使用したいデータを特定し (図 4)、利用を希望するデータの情報を記録したファイル (利用依頼ファイル) をポータルが出力する。この利用依頼ファイルを利用者は管理者に電子メール等の手段で送付する。管理者は受け付けた利用依頼ファイルを確認し (図 5)、利用を許可する場合に、ポータルに対して利用申請ファイルを指定することで、ポータルでは利用条件に一致するデータを自動で検索し、要求されたマスキングなどの加工をした状態のデータを作成する。管理者は作成したデータの情報を利用者に回答



することで、利用者は必要なデータが準備されたことが分かり、利用が可能となる。

【機能 4】データ提供形態 ポータルはバックアップソフトウェアと連携し様々なバックアップ単位、例えば、データベース単位、データベースのインスタンス全体、データベースが稼動する VM 単位など様々な設定でデータ提供が可能である。VM 単位の場合は VM 管理ソフトウェアとも連携する。VM については、データ準備時に OS、アプリケーション(データベース)をインストールした VM を自動作成し、バックアップファイルか

ら復元したデータベースを接続する方法と、事前に OS、アプリケーションを VM に準備して、データベースを接続する方法をサポートすることで複数の管理ソフトウェアを連携した複雑な処理を自動化する。これらの自動化はジョブスケジュールソフトウェアにて実装する。

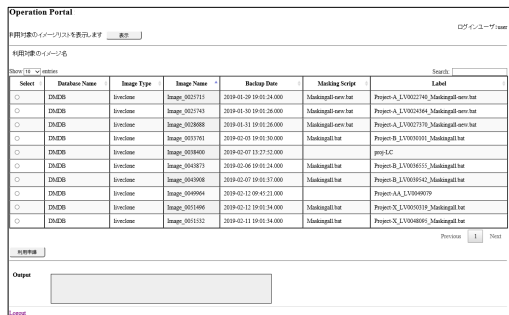


図 4 データ活用ポータルのプロトタイプにおけるデータ利用者向けポータル入力画面

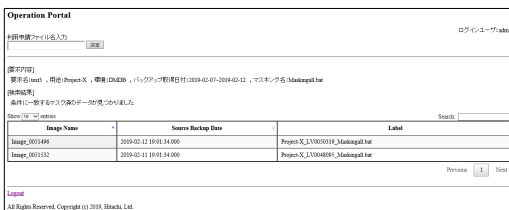


図 5 データ活用ポータルのプロトタイプにおけるデータ管理者向けポータル入力画面

4 データ活用ポータル導入効果の評価

4.1 開発環境提供時間の評価

データ活用ポータルの導入によりバックアップデータを活用した開発環境提供までの時間を以下の 2 つの作業 (1)(2) に分解し、机上計算にて評価した。

(1) 利用可能なデータの検索と利用環境準備作業 ポータル導入前は、利用者は要求を作成し、管理者に送る作業を対象とする。管理者は、利用者の要求を受けて作業する。ポータルでは、対象の作業者を利用者とし、利用者が利用可能なデータを検索、要求して、提供を受ける作業となる。

(2) マスキング済みデータの準備 マスキング済みデータの作成作業対象の作業者は管理者である。

評価尺度として、オペレーション数と作業時間でポータル導入前のバックアップ管理ソフトウェアのみを使用する場合とポータル導入後の場合について、作業時間を算出した。オペレーション数は作業フローの画面操作数から抽出した。作業時間は、処理の開始から完了までの時間を概算した。内訳は作業者が考える時間、作業者が操作する時間、システムでの処理時間とする。操作する時間は 1 オペレーションあたり 0.5 分として算出した。評価結果を表 1 に示す。ポータルの導入により、(1) の作業について、管理者の作業はなしとなる。利用者の作業時間は、ポータル導入前と比較して 2/3 に短縮できる。(2) の作業については、管理者の作業時間を約半分に短縮できる。よって、作業時間を合計すると、利用者がデータを利用可能となるまでの時間は、ポータル導入によって約 1/3 に短縮できることを確認した。

表 1 環境提供処理時間評価

| 処理 | バックアップソフトウェアのみ | | バックアップソフトウェア+ポータル | |
|---------------|----------------|-----------------|-------------------|----------------|
| | 利用者 | 管理者 | 利用者 | 管理者 |
| 環境の検索と準備 | 6 操作 約 7 分 | 17 操作 約 30 分 | 6 操作 約 5 分 | 0 操作 0 分 |
| マスキング済みデータの準備 | － | 16 操作 約 5 分 | － | 2 操作 約 18 分 |

4.2 開発環境イメージファイルの再利用と定期的な削除作業に関する作業時間の評価

「イメージ再利用」と「定期的な掃除作業」というユースケースに対してデータ活用ポータル適用時の効果を確認することを目的に、それぞれのケースの実行手順を検討し、データ活用ポータルの適用効果を机上で評価した。これらのユースケースの実行手順検討では、作業手順を詳細化し、実際に行う作業を明らかにする。評価の前提条件として、ポータル適用前の既存の開発環境 (図 1) では、開発対象データベースを稼働させる VM は事前に構築済みであると仮定した。さらに、「イメージファイル再利用」の前提としては、再利用するバックアップイメージファイルのコピーを活用し、開発環境で活用されるデータベースに登録されたデータは、マスキング処理が施されているため、再利用時に再度マスキング処理を実施しないことを仮定した。「定期的な削除作業」の前提としては、掃除作業の対象を VM および、全てのイメージファイル (データベースのバックアップイメージファイル、それをコピーしたファイル等) とした。VM イメージファイルやデータベースのバックアップイメージファイルには、事前に保存期間が設けられていることを仮定した。

机上にて各ユースケースの操作を評価したところ、「イメージ再利用」で 15 ステップの処理を実施し、「定期的な削除作業」で 16 ステップの処理を実施することが分かった。評価結果の詳細については本論文の付録に記載する。

「イメージファイル再利用」にデータ活用ポータルを適用した場合には、要求に合わせた一覧表示、統合管理による人手作業の短縮、および、処理の自動化による確認作業の割愛等により 8 ステップの処理で実施できることが分かった。「定期的な削除作業」にデータ活用ポータルを適用した場合には、利用者による依頼作業の割愛、VM イメージファイル情報とデータベースのバックアップイメージファイル情報をまとめて自動的な抽出、削除対象を反映させた一覧の表示および、保存期間の自動延長等により 11 ステップの処理で実施できることが分かった。上記検討結果からデータ活用ポータルの適用により「イメージファイル再利用」と「定期的な削除作業」で約 40 % の作業工程数の削減が期待できることが明らかになった。この結果は、単純なステップ数で効果換算しているが、複数の管理ツールの利用および、人手による操作・判断等によるオーバーヘッドを解消が可能であるため、さらなる効果を期待できる。

5 考 察

本研究は、システム障害への対策で取得しているバックアップデータを他の目的に活用するシステムを対象にしており、データ活用ポータルにて提供する活用データに関する情報を一元管理と複数の管理ソフトウェアを使用する操作を自動化によってデータが利用可能になるまでの準備期間を削減を実現した。この技術は、データ管理者に対して、データの二次利用に関わる作業をポータル画面上の操作に抽象化することで、個々のバックアップ管理ソフトウェアや VM 管理ソフトウェア等の管理ソフトウェアの操作の習熟を不要にすることができる。つまり、データ管理者が個々の管理ソフトウェアの学習することを不要にし、学習にかかるコストを削減できる。

一般に、サービス提供中の業務システムのデータベースから他の目的のためにデータを取得する場合、例えば、稼働中の決済系システムに対して、テンポラリに決済とは別用途(例えば分析やシステム開発用)でデータを取得するために大量のレコードの読み出しを行うと、決済処理の応答時間が遅くなりサービスレベルが低下する可能性がある。そのため、サービスレベルの低下が許容されない状況では、サービス停止期間中にデータを抽出するか、システムの障害対策のために取得しているバックアップデータを流用することが考えられる。本研究では業務システムの運用への影響が少ない後者のアプローチで検討を進めた。

一方、OLTP (Online Transaction Processing) システムと OLAP (Online Analytical Processing) システムを共存できるデータベース管理システム (DBMS (Database Management System)) [11, 12] が開発されており、将来的にはこの技術を活用した DBMS を活用することで目的外利用データの抽出に性能面での問題は解決する可能性がある。ただし、システム開発時の要件として決済等の通常業務処理と分析データの抽出処理を考慮したハードウェアリソース規模の見積もりや、実行する処理のスケジューリング、それらを考慮した上での資産コスト、運用コストが投資に見合うかの検討が適用の前提になる。本研究では OLTP と OLAP の両方を同一データベースインスタンスで対応可能な技術を搭載していない DBMS も対象とするため検討対象外とした。

データの再利用を組織間で行っている事例として、疾病に関わる希少なデータを他の医療機関に提供しと共有しあうことで互いに貢献をしあう検討が進んでいる [13]。このようなケースではデータの利用者が期待するデータの属性や内容に大きな乖離はないと思われる。しかし、一般的な業務システムの場合、目的外の使用に耐えられるデータであるかどうか、例えば商品売買の取引データを入手しても、支払いの合計値のみしか記録されていない場合は、各顧客がどんなものを購入しているか分からないため顧客単位でのマーケティングには向いていない。つまり、データを目的外利用で流通させる際にはデータの内容まで踏み込んだ情報提供が必要である。今後、そのような情報もポータルで管理できるようにすることでデータ活用ポータル

の適用先が広がると考えられる。そのためにはセキュリティやデータアクセス権限、アクセスコントロール、データの品質、ライフサイクル管理等を統制したデータガバナンス [14] の体系化とその適切な運用方法の確立が必須と考えられる。

6 ま と め

バックアップデータをシステム開発におけるテストデータに活用する等の本来の目的とは異なる用途への活用を迅速に行うためのデータ活用ポータルの提案をした。実際のシステム開発事例をもとにユースケースを検討し、データの利用許可の承認や要求データの特定制、複数の管理ソフトウェアを人的作業で連携させていることによるデータ提供が迅速に行われない問題や誤操作リスクが増大することを明らかにした。そして、この問題の解決のためにデータの一元管理と複数の管理ソフトウェアを自動的に連携した処理を行うデータ活用ポータルを提案し、机上にてその効果を検証した。その結果、導入前と比べてデータが利用可能になるまでの時間は約 1/3 に削減可能であり、開発環境で活用したデータの再利用や不要なデータ削除の工程数は約 40 %削減できることを示した。さらに、ポータルの導入によりデータ管理者が個々の管理ソフトウェアの操作に習熟する必要がなくなり、データ管理者の学習にかかるコストを抑える効果もあることを検討した。

今後は、業務システムデータを分析など他の用途で流通させるために必要なデータ管理手法の一般化を行い、データ主導によるビジネス改革への貢献と、データガバナンス技術の体系化に貢献していきたい。

文 献

- [1] 保田勝通. ソフトウェア工学の現状と動向: テスト・品質保証技術の現状と課題. *情報処理*, 28(7), 1987.
- [2] 西村崇. 記者の目 実データは設計フェーズで役立てよう, 日経クロステック. <https://tech.nikkeibp.co.jp/it/article/Watcher/20120120/379026/>, 2012. Accessed: 2019-12-20.
- [3] 久枝稯. システム導入のためのデータ移行ガイドブッカーコンサルタントが現場で体得したデータ移行のコツ, インプレス R&D, 第 5 章データ移行, 2017.
- [4] Actifio Inc. Test data management for faster application development. <https://www.actifio.com/solutions/use-cases/dev-ops-test-data-management/>. Accessed: 2019-12-18.
- [5] Harvinder Atwal. *Practical DataOps: Delivering Agile Data Science at Scale*, chapter 7, page 182. Apress. <https://doi.org/10.1007/978-1-4842-5104-1>.
- [6] HV Jagadish. Big data and science: Myths and reality. *Big Data Research*, 2(2):49–52, 2015.
- [7] Actifio Inc. Enterprise cloud data management actifio. <https://www.actifio.com/>. Accessed: 2019-12-18.
- [8] Veritas Technologies LLC. Veritas the leader in enterprise data protection. <https://www.veritas.com/>. Accessed: 2019-12-18.
- [9] The leader in cloud data management – veeam software. <https://www.veeam.com/>. Accessed: 2019-12-18.
- [10] Github – ansible-awx awx project. <https://github.com/ansible/awx>. Accessed: 2019-12-18.
- [11] Hasso Plattner. A common database approach for oltp and olap using an in-memory column database. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 1–2, New York, NY, USA, 2009. ACM.

[12] A. Kemper and T. Neumann. Hyper: A hybrid oltp olap main memory database system based on virtual memory snapshots. In *2011 IEEE 27th International Conference on Data Engineering*, pages 195–206, April 2011.

[13] Yoshihiko Furusawa, Izumi Yamaguchi, Naoko Yagishita, Kazumasa Tanzawa, Fumihiko Matsuda, Yoshihisa Yamano, RADDAR-J Research, and Development Group. National platform for rare diseases data registry of japan. *Learning Health Systems*, page e10080, 2019.

[14] Vijay Khatri and Carol V Brown. Designing data governance. *Communications of the ACM*, 53(1):148–152, 2010.

付 録

A 開発環境イメージファイルの再利用と定期的な削除作業
に関する作業時間の評価結果

表 A・1 開発環境イメージファイルの再利用処理時間評価

| Step | 実施内容 | ポータル適用時の効果 |
|------|---|---|
| 1 | バックアップイメージ一覧を取得 | － |
| 2 | 利用要求の選定 | ポータル画面上で利用要求に合 |
| 3 | 利用要求に合わせたイメージの検索 (または、一覧の再表示) | わせた一覧表示により検索手順・時間を短縮 |
| 4 | 再利用対象の VM 情報 (利用 DBMS, VM の利用リソース等) を取得 | 統合管理により VM とイメージ情報を纏めて取得可能なため、人手による各ツール操作手順・時間を短縮 |
| 5 | 再利用対象のイメージ情報 (容量, 利用環境等) を取得 | |
| 6 | 開発環境の VM 側リソース情報取得, または、再利用向け VM の検索 | 統合管理により VM とストレージ情報を纏めて取得・判断可能なため、人手による各ツール操作手順・時間を短縮 |
| 7 | 開発環境のストレージ側リソース情報取得 | |
| 8 | 再利用向け環境の構築可否を判断 | |
| 9 | イメージ再利用向け環境準備 | 環境準備と確認を自動的に実行可能 |
| 10 | イメージ再利用向け環境確認 | |
| 11 | 再利用したいイメージのスナップショットを取得 | スナップショットの取得と確認を自動的に実行可能 |
| 12 | 取得したスナップショットの確認 | |
| 13 | バックアップイメージを開発 DB サーバに接続 | 接続処理と確認を自動的に実行可能 |
| 14 | 開発 DB サーバへの接続状況確認 | |
| 15 | DBMS にアクセスしてイメージ再利用を開始 | － |

表 A・2 開発環境イメージファイルの削除時間評価

| Step | 実施内容 | ポータル適用時の効果 |
|------|--|---|
| 1 | 各 VM の保存期間, 稼動情報をスクリプト等により定期的に取得 | － |
| 2 | VM に接続した (バックアップイメージファイル内のデータベースに登録された) データの内, 保存期間切れデータの有無の確認を依頼 | ポータルで定期的に行うため, 確認依頼は不要 |
| 3 | VM に接続した (バックアップイメージファイル内のデータベースに登録された) データの保存期間の情報を取得し, 保存期間切れデータを抽出 | － |
| 4 | 各 VM の保存期間 (経過時間) や稼動情報から削除対象 VM を抽出※保存期間切れ VM の抽出完了 | Step1 の取得情報から削除対象 VM, および, 本 VM に接続されるデータを自動的に抽出されるため, 抽出作業・時間を短縮 |
| 5 | 削除対象 VM に接続される各 (バックアップイメージファイル内のデータベースに登録された) データを抽出 | |
| 6 | VM に接続した (バックアップイメージファイル内のデータベースに登録された) データの内, 抽出した保存期間切れデータを報告 | ポータルが各 VM 情報の取得, 削除対象 VM の抽出結果を反映させた一覧表示により削除可否の判断時間を短縮 |
| 7 | 保存期間切れ (バックアップイメージファイル内のデータベースに登録された) データをマウントする VM を削除対象として抽出 | |
| 8 | 削除対象 VM (Step5, Step6 で抽出) と削除対象データ (Step6, Step7 で抽出) の削除可否を確認 | |
| 9 | 削除対象 VM の削除が不可な場合, 削除理由が「VM の保存期間切れ」であれば, 削除対象 VM の保存期間を延長 | － |
| 10 | 削除対象 VM の削除が不可な場合, 削除理由が「バックアップデータの保存期間切れ」であれば, 削除対象データの保存期間を延長するように依頼 | 削除不可となった VM 情報を取得し, VM のデータ保管先を選択し, 自動的に保存期間を延長するため, 依頼作業は不要 |
| 11 | 削除対象データの保存期間を延長する依頼を受けた場合, 対象データの保存期間を延長 | |
| 12 | 削除対象 VM の削除が可能な場合, 削除対象 VM に接続したバックアップデータを取り外す | － |
| 13 | 削除対象 VM (Step9, Step10 の VM 以外) を順次削除 | － |
| 14 | 削除対象 VM から取り外したしたデータベースイメージファイルのデータの削除を依頼 | バックアップソフトウェアの提供機能によりデータを自動的に削除されるため, 人手による操作不要 |
| 15 | 開発環境の管理者から削除依頼があったバックアップデータを (一括) 削除 | |
| 16 | 削除対象 VM の整理を一通り完了したことを確認 (VM の定期的な掃除作業完了) | － |