

分析データベースの可用性を向上させる “AnalyticMartレプリケーションサービス”

森川修一*

"AnalyticMart Replication Service" : High Availability for Data Warehouse

Shuichi Morikawa

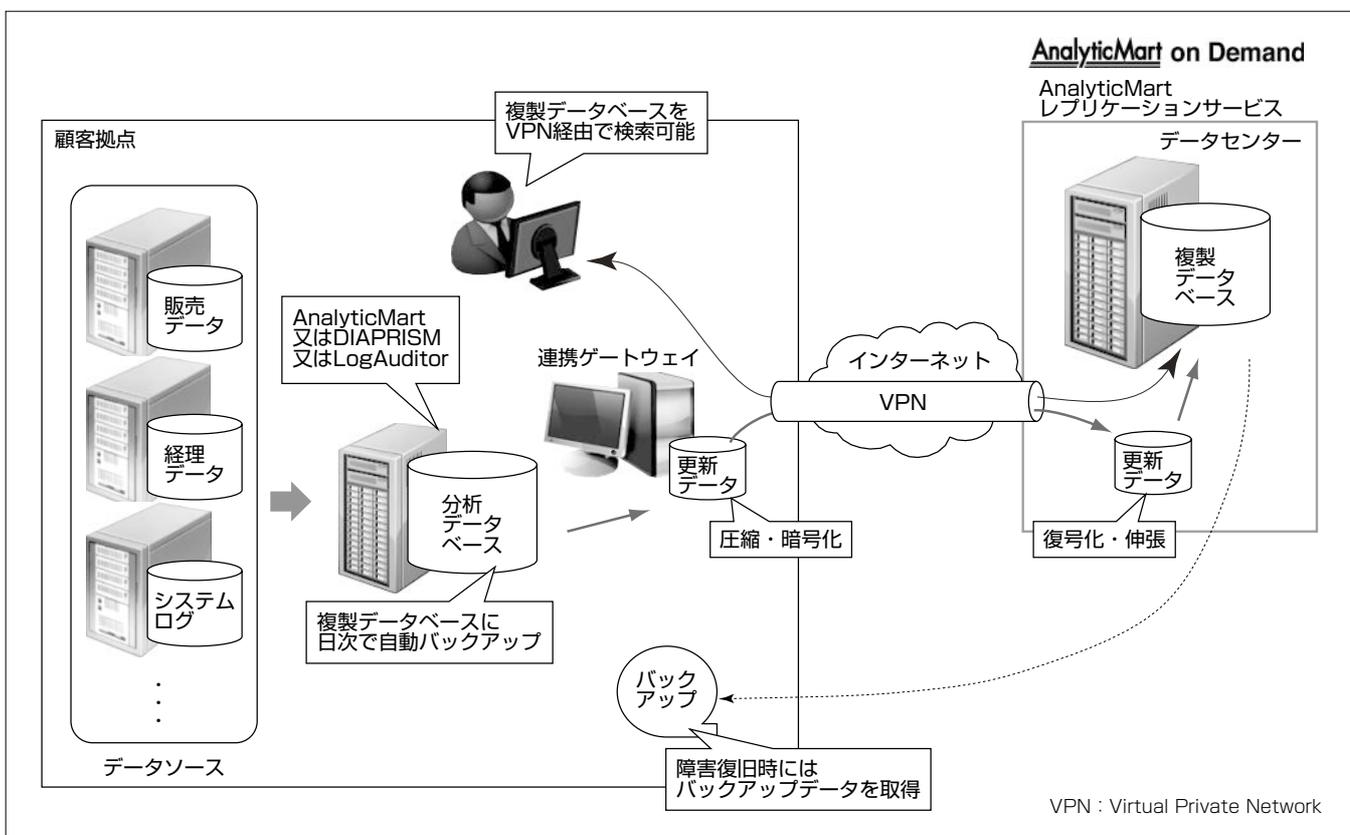
要旨

スマートフォンの普及などによって個人の行動に関わるデータが日々蓄積され、その増加傾向は著しい。データ活用に対する意識も高まり、データ分析を積極的に業務運営に取り入れる企業も増えている。業務におけるデータ分析の必要性が高まるにつれて、分析用として構築・運用するデータベース(分析データベース)に障害が発生した場合の業務への影響も看過できなくなってくる。

三菱電機インフォメーションテクノロジー(株)(MDIT)のデータ分析クラウドサービス“AnalyticMart on Demand”のサービスの1つである“AnalyticMartレプリケーションサービス”では、クラウド上に複製データベースを自動作成することで障害による業務への影響を小さくする。

クラウド上の複製データベースは日次でデータを自動更新して利用可能な状態で常時稼働する。分析データベースの障害発生時には検索対象を複製データベースに切り替えることで、分析データベースの復旧を待たずに業務を再開できる。バックアップは顧客の拠点とは地理的に異なる所在地の複製データベースに保管され、バックアップの破損・不備によるデータ消失を回避できる。復旧では複製データベースからの完全バックアップデータ取得による1回のリストア作業で済み、作業工数を削減できる。

“AnalyticMartレプリケーションサービス”によって障害からの早期の業務再開と復旧時間の短縮が可能になり、分析データベースの可用性を向上させることができる。



“AnalyticMartレプリケーションサービス”の概略

各種データソースから抽出、加工、格納した分析データベースをデータセンターに自動で複製する。分析データベースはMDITの“AnalyticMart” “DIAPRISM” “LogAuditor”に対応する。携帯ゲートウェイで更新データを日次で収集、圧縮、暗号化してデータセンターに送信し、複製データベースに反映する。障害時は複製データベースに切り替えて業務を再開、バックアップデータを複製データベースから取得してデータを復元できる。

1. ま え が き

ポイントカードや電子マネーの定着、スマートフォンの普及、SNS(Social Networking Service)の利用者拡大等によって、個人の行動に関わるデータが日々生成、蓄積されてきている。近年のビッグデータの流行や行政機関によるオープンデータの推進、IoT(Internet of Things)に向けた企業の取組み等、世の中で生成・蓄積されるデータの増加は加速する一方である。

企業では、顧客及び見込み顧客に関して広範囲で細かい粒度、頻度のデータを収集できる環境が整いつつある。収集したデータを利用して、これまでのKKD(勘・経験・度胸)に頼った意思決定をデータ分析に基づいた判断で行う企業も増えてきている。

業務遂行でのデータ分析の必要性が高まるにつれ、データ分析システムに障害が発生した場合の業務への影響は許容できない域に近づいている。特に分析用に様々なデータソースからデータを抽出、加工して蓄積するデータベース(分析データベース)では稼働停止による業務への影響は大きい。企業では必要最小限の対策として分析データベースのバックアップを実施している場合が多いが、不幸にも障害が発生した際に分析データベースの復旧に難航することも珍しくない。

本稿では、分析データベースの障害発生時におけるバックアップからの復旧に関わる課題を述べるとともに、その解決策となる“AnalyticMartレプリケーションサービス”(略称“AnalyticMart/RL”)について述べる。

2. バックアップによる障害復旧の課題

一般的なデータベースの障害対策には、定期的なバックアップ運用がある。日次や週次といった周期に従ってデータの複製(バックアップ)を採取して保管する。データベースの障害発生時には保管してあるバックアップデータからデータを復元することでデータベースの復旧を行う。このようなバックアップを用いた障害復旧には次に述べる3つの課題(図1)がある。

2.1 復旧までの業務停止時間

バックアップによる障害復旧では、データベースの障害で業務が停止した場合、データを復元してデータベースの運用を再開するまでの間、業務は停止したままとなる。基

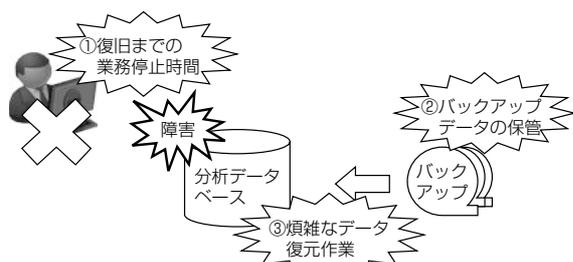


図1. 障害復旧に関わる課題

幹系のトランザクション処理などでは暫定的に直近のデータを復元して業務を再開し、段階的に復旧を進めるといった対応もできる。しかし分析データベースでは、過去の履歴を中心とした蓄積データが処理の対象となるため、バックアップからのデータ復元を待つ必要がある。一般的に分析データベースは運用に伴ってデータ量が増大していくため、運用期間が長くなるにつれてデータ復元に必要な時間も増加し、業務停止時間も長くなる。

データベース二重化などの冗長構成によって障害発生時にも業務を継続するといった対策もあるが、中堅・中小企業では費用などの点から導入が難しいのが実情である。

2.2 バックアップデータの保管

通常、バックアップデータは自社システム内のストレージ機器や磁気テープ等の媒体に記録し保管する。ストレージ機器に保管する場合、機器の設置や運用管理が必要となる。媒体に保管する場合は、媒体の破損や劣化を予防し、盗難や不正アクセス等の物理セキュリティに配慮した保管場所を選定する必要がある。

バックアップ用のストレージ機器の設置場所又は媒体の保管場所がデータベースと同一拠点内にある場合、自然災害などの地理的影響範囲を持つ要因による障害では、データベースと同様にストレージ機器や媒体に記録されたバックアップデータも被害を受ける可能性がある。BCP(Business Continuity Plan)対策の観点からはバックアップデータは地理的に異なる拠点で保管することが望ましい。

またバックアップ運用が計画通り実行され、バックアップデータの保管に問題がないにも関わらず、いざ復旧の際に論理的なデータの欠損や不足によってデータの復元に失敗するといった事例も発生している。バックアップデータの整合性はシステム構築時に検証されるだけで、バックアップ運用の開始後は検証されることがほばないためである。

2.3 煩雑なデータ復元作業

通常バックアップは採取時間と復元時間の効率化の兼ね合いから、完全バックアップ(すべてのデータの複製)、差分バックアップ(完全バックアップ以降の変更データの複製)、増分バックアップ(前回バックアップ以降の変更データの複製)の組合せで運用することが多い。例えば月次で完全バックアップ、週次で差分バックアップ、日次で増分バックアップを採取するといった運用である。このような運用でデータを復元する場合、完全バックアップデータのリストア1回、差分バックアップデータのリストア1回、増分バックアップデータのリストアをデータの個数分実行することになる。

障害発生時にはバックアップの採取周期を確認し、複数回のリストアを順番に実行する必要がある。リストアを定期的実施することはないため作業の習熟は期待できず、実際の復元作業では過度な緊張などから作業手順の抜けや誤りが発生する危険性をはらんでいる。

3. AnalyticMart/RL

3.1 AnalyticMart on Demand

AnalyticMart/RLはMDITのデータ分析クラウドサービス“AnalyticMart on Demand”が提供するサービスの1つである。AnalyticMart on Demandのサービスを表1に示す。

AnalyticMart on DemandはMDITのデータ分析フレームワーク“AnalyticMart”⁽¹⁾⁽²⁾の構造化データ用データベース(AQL)を利用したサービスをクラウド型で提供する。サービス化によってシステムへの設備投資や運用人件費を削減し、初期費を抑えたスモールスタートが可能となる。

顧客の拠点とクラウドのデータセンターとの間のネットワークには、インターネットでの盗聴やセッション乗っ取り等の防止対策が施された三菱電機情報ネットワーク(株)のセキュアネットワークサービスを利用することで安全性を確保している。

3.2 システム

AnalyticMart/RLでは顧客拠点で稼働する分析データベースの複製データベースをデータセンターに自動作成する。複製が可能な分析データベースは表2に示すMDIT製品である。

AnalyticMart/RLのシステム構成を図2に示す。分析データベースは顧客の拠点にあるAQLサーバで運用される。AnalyticMart/RLではAQLサーバの分析データベース①をデータセンターの仮想マシン上で動作する複製データベース②に自動複製する。仮想マシンは顧客ごとに専用環境を用意するため、複製データベースは他の顧客から完全に分離、保護される。

顧客がAnalyticMart/RLを利用するにはインターネット回線のほかに連携ゲートウェイ用のパソコンを用意する必要がある。連携ゲートウェイでは分析データベースの複製などを行うレプリケーション機能④と、各種設定や運用監視を行うAQLレプリケーションマネージャー⑤を実行す

表1. AnalyticMart on Demandのサービス

サービス	内容
AnalyticMartプラットフォームサービス AnalyticMart/PF	データ分析システムの基盤をクラウド上に提供するサービス
AnalyticMartレプリケーションサービス AnalyticMart/RL	顧客の拠点にある分析データベースの複製をクラウド上に作成するサービス
AnalyticMartレポートサービス AnalyticMart/AG	省エネ法の“定期報告書”で必要な事業所ごとのエネルギー使用量の集計を行う受託集計サービス

表2. 対応するデータベース

製品名	コンポーネント
データ分析フレームワーク AnalyticMart	AQL (version 4.0.0以降)
	AQL Express Class 1 (version 4.0.0以降)
	AQL Express Class 2 (version 4.0.0以降)
	AQL Express Class 3 (version 4.0.0以降)
データ分析プラットフォーム DIAPRISM	DIAPRISM/AQL (N00版以降)
統合ログ管理ソリューション LogAuditor	LogAuditor/AQL (N00版以降)

る。運用中の分析データベースにAnalyticMart/RLを導入する際に既存システムへの影響を低減できるようAQLサーバとは別に連携ゲートウェイが必要な構成としている。レプリケーション機能④では顧客専用仮想マシンで稼働するレプリケーション機能(サーバ)③と連携し、ファイル駆動⁽³⁾による更新データの反映やリストアデータの取得を行う。顧客拠点と顧客専用仮想マシン間の通信ではデータ圧縮及び暗号化を行っている。AQLレプリケーションマネージャー⑤ではWebインタフェースによってレプリケーション機能④及びレプリケーション機能(サーバ)③の動作設定や運用状態の確認を行うことができる。

顧客拠点から複製データベースにアクセスするための設備及びアプリケーション⑥については、顧客が分析データベースの利用時に使用するものがそのまま使用できる。

AnalyticMart/RLの利用には顧客拠点での環境構築が必要となる。契約後から利用開始までに顧客が実施する主な作業は次のとおりである。

- (1) VPNルータが送付されてくるので、VPNルータを設置し、顧客拠点とデータセンターの顧客専用仮想マシンの接続を確立する。
- (2) 連携ゲートウェイに必要なソフトウェア類一式を顧客専用仮想マシンからダウンロードする。
- (3) 連携ゲートウェイに(2)でダウンロードしたソフトウェアをインストールする。
- (4) 連携ゲートウェイから分析データベースのバックアップを収集できるようにAQLサーバの環境設定を行う。

3.3 機能

AnalyticMart/RLではレプリケーション、リストア、データ検索の3つの機能を提供する。

(1) レプリケーション機能

顧客専用仮想マシン上に分析データベースを複製し、更

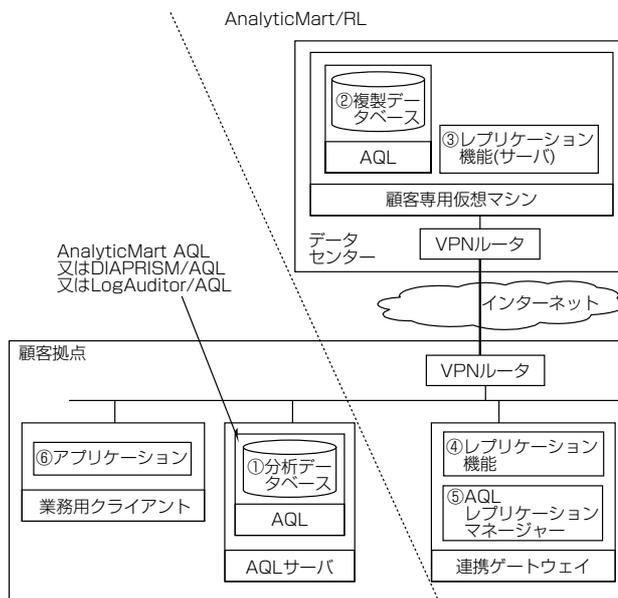


図2. システム構成

新データを日次で同期する処理を自動実行する。レプリケーションの運用状況はAQLレプリケーションマネージャーからWebブラウザを用いて確認できる。日次処理の完了通知/異常通知を電子メールで送信することも可能である。

(2) リストア機能

複製データベースから完全バックアップデータを採取し、連携ゲートウェイにダウンロードする。連携ゲートウェイから分析データベースへのリストアの自動実行は機能として提供していないため、別途顧客が実施する必要がある。

(3) データ検索機能

顧客拠点のクライアントから複製データベースに対してデータ検索を行うことができる。実行できる操作をデータ検索だけに限定しているため、顧客の誤操作によって分析データベースとの同期が損なわれる心配がない。

3.4 特長

2章で述べたバックアップにおける障害復旧の課題に対する解決策という観点からAnalyticMart/RLの特長(図3)を述べる。

3.4.1 複製データベースによる業務再開

顧客専用仮想マシン上の複製データベースには分析データベースの内容が日次で反映され、前日までのデータが利用可能な状態で稼働している。障害発生によって分析データベースが運用を停止したとき、顧客は検索対象を複製データベースに切り替えることによって業務を再開できる。複製データベースでは障害発生当日のデータは利用できないため再開可能な業務が限定される場合もあるが、部分的な再開であっても障害による業務への影響を緩和できる。

分析データベースの復旧を待たずに業務を再開できることから、蓄積データ量の増加によるデータ復元時間の延長が業務再開時間に影響を与えない。障害発生から一定時間での業務再開が可能なることから、分析データベースの可用性を向上させることができる。

また複製データベースは、障害復旧時に限らず利用できるため、日常業務のデータ検索操作の一部を複製データベースで代行することによって負荷分散を図るといった用途にも活用できる。

3.4.2 クラウドによるバックアップ

データセンターの顧客専用仮想マシンで稼働する複製データベースにバックアップデータが格納されるため、顧客はバックアップ用の機器運用や媒体保管等を行う必要がない。物理セキュリティ対策についてもAnalyticMart/RLの運用の一環として提供されるため、顧客自身による対策は不要である。顧客拠点とバックアップデータの所在が地理的に分かれることから、顧客拠点での事故や火災による障害の影響を受けにくい。

バックアップが複製データベースとして稼働していることから、顧客はデータの保管状態や論理的な整合性を容易

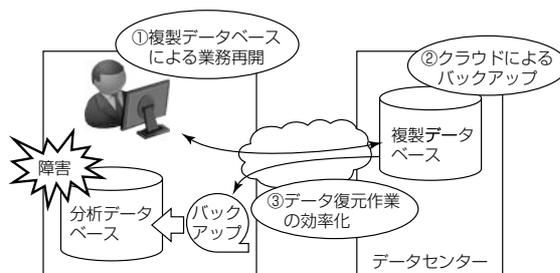


図3. AnalyticMart/RLの特長

に確認できる。障害復旧の際にバックアップデータの状態が原因でリストア不能になるといった事態を回避できることで信頼性が向上する。

3.4.3 データ復元作業の効率化

データ復元で使用するバックアップデータは普段は複製データベースとして存在しており、リストア可能な形態ではない。障害などによってバックアップデータが必要になった時点で顧客が複製データベースから取得する。そのため一般的なバックアップのように完全/差分/増分によるバックアップデータの組合せではなく、常に完全バックアップデータだけとなる。データ復元作業は完全バックアップデータからの一回のリストアで終了するため、作業回数が低減されて復元時間が短縮できる。データ復元の作業工数が削減できることから保守性が向上する。

また、複製データベースの作成・更新で使用する更新データの採取では、初回の分析データベースからの完全バックアップ以降は常に増分バックアップが更新データとなるため、日次で実行する更新データの採取時間も短縮できる。

4. むすび

AnalyticMartレプリケーションサービスでは、顧客拠点の分析データベースに障害が発生したとき、クラウド上の複製データベースで業務を再開することによって可用性を向上させることができる。障害復旧ではバックアップデータを複製データベースから取得することで確実な復旧と作業工数の削減が可能となる。今後はデータセンターの地理的な分散などによって、可用性や保全性の更なる向上を図っていく予定である。

参考文献

- (1) 小出健太, ほか: 情報セキュリティを支えるデータ分析フレームワーク“AnalyticMart”, 三菱電機技報, **87**, No.7, 405~408 (2013)
- (2) 和田貴成, ほか: 統合ログ管理ソリューション“AnalyticMart for LogAuditor”, 三菱電機技報, **86**, No.7, 391~394 (2012)
- (3) 山岸義徳, ほか: ファイル駆動によるデータレプリケーション方式, 電子情報通信学会2014年総合大会, D-4-1 (2014)