

ZL Unified Archive

ビッグデータと巨大サイロに伴う深刻な問題

ZL TECHNOLOGIES | ホワイトペーパー

ビッグデータはデータ管理において益々深刻な問題となりつつあります。多くの場合、ビッグデータは、企業全体の至る所に無秩序に散在し、データサイロがあちこちに作成されるため、その管理はとりわけ困難となります。諸々の規制とeディスカバリー制度の制定に伴い、企業はこの無秩序に増大を続けるデータの管理を余儀なくされました。このホワイトペーパーでは、ビッグデータに伴う問題の概要を述べ、その診断方法と解決策を説明します。さらに、次世代に向けてのビッグデータ管理のイノベーションについても触れます。

ビッグデータの簡単な歴史 - 構造化データ/非構造化データ

ビッグデータは害悪と見なされがちですが、大きなメリットがあります。ただし、そのメリットを認識するためには、まず「構造化データ」と「非構造化データ」の違いを理解する必要があります。

構造化データは、コンピューターによりコンピューターのために生成されるフォーマット化された情報です。これには、ERP システム、データウェアハウス、トランザクションなどのデータが含まれます。1990 年代に入るまでは、電子データのほとんどが構造化データでした。

これとは対照的に、非構造化データは主に人間により人間のために作成されるフォーマット化されていないテキストデータであり、ワードプロセッシングの普及に伴いその量は増大しました。2000 年までに電子メールやファイルなどのデジタルデータが爆発的に普及し、データのストレージをめぐる既に深刻な問題が生じるようになりました。この時期に、企業による不正行為と訴訟問題が立て続けに起こり、その結果、電子保持情報 (ESI) の管理、保全、開示を義務付ける一連の政府規制と法規定が新たに制定されました。これは企業にとって新たな課題となり、その対策として以下のような一時的ソリューションの導入が進みました。

電子メールとファイルのストレージ最適化

高コストなプロダクションサーバー上のメールとファイルの保存に使用される記憶域容量を削減することは、IT 部門にとって常に困難な課題となっています。2000 年には、サイズの大きなファイルと添付ファイルをプロダクションサーバーから低コストのアーカイブにオフロードし、元のドキュメントに小さなリンクまたはスタブを残すという解決策が考案されました。エンドユーザーがスタブをクリックすると、ドキュメントが元の形式に復元されます。この方法では、メールとファイルの保存に使用される記憶域容量が最大 90% まで削減可能となり、ストレージコストと管理コストを大幅に節約できます。

コンプライアンス

エンロン社を始めとする一連の企業スキャンダルの結果、規制機関により多数の法規制 (SEC 17a-4、NASD 3010、FERC 717 など) が制定され、メールとファイルの保持・開示要件が強化されました。

すぐに金融業界だけでなくヘルスケアやその他の業界にも同様の規制が施行されるようになりました。今日大企業は、複数の法規制によりデータ保持が義務付けられており、効果的なセルフガバナンス体制の必要性が高まっています。

e ディスカバリー

2006 年 12 月、米国連邦民事訴訟規則（「FRCP」、Federal Rules of Civil Procedure）により電子的な証拠データの取り扱いに関する原則が定められました。これを機に、メールやファイルなどの非構造化コンテンツの管理の重要性が高まりました。その後数年の間に、電子情報開示の必要性に伴い、e ディスカバリープロセスとパフォーマンスを向上するための新技術が開発されるようになりました。

レコード管理

2000 年代後半には、当時提供されていた簡素なツールでは、企業が必要とする電子レコード管理に対応できなくなりました。一般的には、エンタープライズコンテンツ管理 (ECM) プラットフォームに簡易ツールを追加することで、レコード管理対策が行われました。作業を簡略化するためには、保持対象となる重要ドキュメントをエンドユーザーが手作業で分類しなければなりませんでしたが、電子メールとファイルは爆発的な増加を続け、手作業のプロセスでは追いつかなくなりました。そこで、すべての電子ドキュメントを迅速かつ正確に分類できる技術が採用されるようになりました。

過去 10 年の間に、必要に応じて一時的な各種ソリューションとアプリケーションが断片的に開発されてきました。そのため、それぞれが独立して機能するポイントプロダクトとして進化し、結果的に相互連携を欠いた個別のデータサイロが生成されることになりました。データサイロでは、その性質上データ分離の問題が生じやすい傾向にあります。大手ベンダーは、これらの単独ソリューションを組み込んだ、より包括的ソリューションの開発を試みましたが、依然としてサイロの寄せ集めに過ぎませんでした。本来、データ管理の問題を解決するために開発されたサイロ構造ですが、さらに多くの問題を引き起こしているのが現状です。

データサイロの 3 つのデメリット

非構造化ビッグデータを管理するために多くの企業がポイントプロダクトを導入しましたが、サイロ構造には以下の 3 つのデメリットがあります。

1. サイロごとに重複するデータコピー。

2. 一貫性のない検索: 各サイロで異なる検索エンジンと検索機能が使用されるため検索結果に一貫性がありません。
同じデータセットに対して同じ検索コマンドを実行しても、サイロごとに異なる検索結果が返されます。
3. 連携性のないデータ保持期間: 各サイロのデータ保持機能が異なるため、サイロ間の連携に欠けます。同じドキュメントに対して、サイロごとに異なる保持期間が適用されます。

これらのデメリットのため、サイロ型ソリューションではデータ管理が十全に機能しません。そのため、サイロ型ソリューションを採用している企業の弁護士が、法廷において提出した証拠が完全かつ有意なものであると宣誓する場合、確信というよりも希望的観測に基づく発言と考えるのが妥当です。

サイロのデメリットを克服するには、単一のシステムにすべてのデータとアプリケーションを統合し、1つのデータコピー、1つの検索方法、1箇所でのデータ保持ポリシーの一元管理方法が使用されるようにする必要があります。総合的データ管理を行うには、これ以外に方法はありません。

買い手側の注意事項

多くのベンダーが、サイロ構造に関するユーザーの懸念に対応し始めました。そのため、1つの完全なシステムを構成する相互に連携した“統合型”ソリューションという概念がマーケティングにおいて重視されるようになりました。統合型アプリケーション間では、プログラムのインターフェイスを介してデータのやり取りが可能ですが、これは水が入った2つのコップを1つのストローでつなげたような構造と言えます。ただしこれには致命的な欠点があり、非構造化ビッグデータのサイズが膨大な場合、このインターフェイスは機能しません。これは、2つの給水塔を細いストローでつなげたようなものです。

また、単なる統合型ソリューションを“ユニファイド型”ソリューションとして宣伝しているベンダーもありますが、真のユニファイドシステムは、次のような特性をすべて備えていることが条件となります。

1. すべてのアプリケーション機能と特性を備えたシングルシステム
2. 単一コードベース
3. すべてのアプリケーションの単一のバックアップとリストア
4. すべてのアプリケーションを対象とした単一のユーザープロフィール
5. すべてのアプリケーションで共有される単一のデータスキーマ

今後の展望

今日、非構造化データはデータ全体の約 80% を占めており、電子メールやソーシャルメディア、マルチメディアなどの通信ツールの普及に伴い急速な増大を続けています。企業はコンプライアンス、e ディスカバリー、レコード管理のため、既に蓄積された膨大な非構造化データを管理する必要性に迫られています。これを機に、単に防衛的な対策を立てるだけでなく、コーポレートメモリーを構築することで、この膨大なデータを活用し戦略的優位性を築くことが可能となります。

Corporate eMemory™ の活用

メール、ファイル、IM、ソーシャルメディアなどの非構造化データに共通する基本的な特性は、人間により人間のために作成されたデータであるという点です。実際、企業における非構造化データは、過去と現在の全従業員の作業の成果物です。このデータの宝庫の活用、管理、マイニングを可能にするのが Corporate eMemory™ 機能です。この機能は、情報技術を活用して戦略上の優位性を築く上で多大な効果をもたらします。

Corporate eMemory™ の活用例を以下に紹介します。

- 知識管理: 特定の業務や分野の専門家を、退職者や元従業員も含めて見つけます。専門家データベースの設定と管理は不要です。
- ガバナンス: 問題のある動作やコンプライアンスの違反行為を早期に発見します。
- CRM (カスタマーリレーションシップマネージメント): コミュニケーションに関するセンチメント分析を実行し、最適なサービスの提供を図ります。
- 研究開発の共有: アイディアを共有し、重複する研究活動に費やされるリソースの無駄を省きます。
- 内部調査: 企業の隅々まで詳細に調査し、迅速に結果を取得します。

上記の例は、非構造化ビッグデータを Corporate eMemory™ へと変換し、戦略上の価値を引き出すための活用例のわずか一部に過ぎません。

企業にとっての Corporate eMemory™ は、人間にとっての記憶 (メモリー) に匹敵するほど、かけがえない価値あるものとなります。

ZL Technologies, Inc. について

ZL Technologies, Inc. は、大企業が e ディスカバリー、レコード管理、規制コンプライアンス、コーポレートガバナンス、ストレージ管理のためのアーカイブニーズに対応するため、電子メールやファイルなどの非構造化コンテンツ全体を管理するためのツール、Unified Archive® ソフトウェア (“ZL UA”) を提供しています。このソフトウェアを通じて、非構造化コンテンツを企業の Corporate eMemory™ に変換し、企業全体で知識、専門技術、履歴などを共有することで競争上の優位性を築くことも可能になります。ZL UA は、比類ない完全に統合されたユニファイドアーキテクチャを特長とし、すべてのアプリケーションと膨大な数のドキュメントを1つのプラットフォームに統合します。その結果、運営コストと訴訟リスクを増大させる主な原因となっている分断されたデータサイロを排除できます。Global 500 企業を顧客とし、Oracle、PriceWaterhouseCoopers、SunGard などの大手企業を戦略的パートナーとする ZL は、その確かな実績の下に、非構造化ビッグデータを活用して戦略上の優位性を確立する技術を提供するテクノロジーリーダーとしての地位を築いています。詳しくは、www.zlti.com をご覧ください。