


VERITAS™

DATA GENOMICS INDEX

非構造化データの専門家によるストレージ環境の
データ構成調査レポート

2016



Data Genomics プロジェクトは、データの管理方法に対する私たちの考え方を考えるために立ち上げられた新しい試みです。ベリタスは志を同じくするデータサイエンティストや業界の専門家、オピニオンリーダーたる有識者を集め日常的に作成、保存、管理されている非構造化データの真の姿をとらえる取り組みを立ち上げました。

本プロジェクトの最初の成果物としてお届けするのが、この『Data Genomics Index』です。これはストレージ環境のデータ構成内容についてまとめた初の調査レポートです。

ベリタスは Fortune 500 社のうち 86% にあたる企業をお客様に持ち、エクサバイト単位のデータのバックアップやアーカイブ、分析を請け負っており、企業環境の本質を示す情報を集めることができる特異な立場にあります。今回、ファイルの属性を示すメタデータに注目し、お客様のご協力のもと、これらを収集・分析することで実際の環境がどのようなデータで占められているのかが明確になりました。

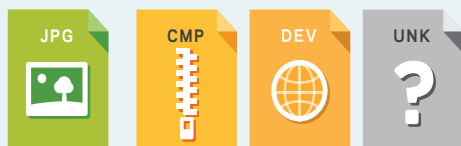
現在のデータ構成内容



41%

のファイルは過去

3年間 更新されていない



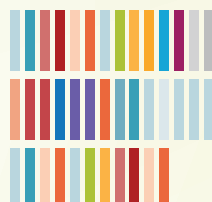
最も多くの容量を占めるファイルは
分類が難しい特殊なタイプである



古い形式の「office」ファイルが
ストレージ容量に負荷を与えている



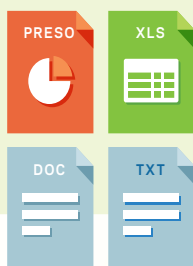
所有者不明データ
(Orphaned Data/
オーファンドデータ)



所有者不明データ(Orphaned
Data/オーファンドデータ)は特
に陳腐化が進み管理負荷増大の
原因となっている

これらをアーカイブすることで

ストレージ費用を削減



50%

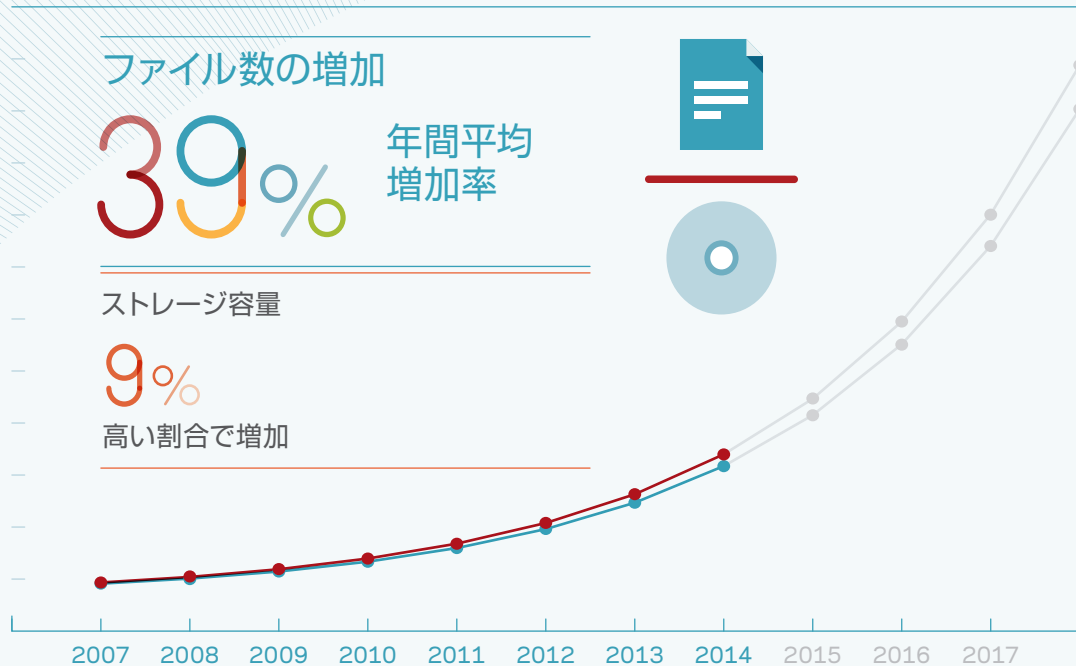
環境負荷の高いファイルを優
先的にアーカイブしたり削除
することで保管コストを削減
することが可能である



新生ベリタスの Data Genomics Index

2015 年ベリタスは、お客様企業の多くの非構造化データ環境から数百億におよぶファイルを収集し、それらを属性とともに分析することでストレージ環境がどのようなファイルで構成されているのかを調査しました。

分析では 8,000 個以上の一般的なファイルの拡張子でタイプを特定し、分類しました。こうして得た情報はそれぞれのお客様が持つ非構造化データの特徴を示しています。



データの爆発的増加

過去7年間を平均すると、ファイル数は前年同期比 39.2481189% の割合で増え続けています。しかしながら図で示すように実際にはこの割合が年々増えてきており、傾向としては指数関数的な増加が予測されます。

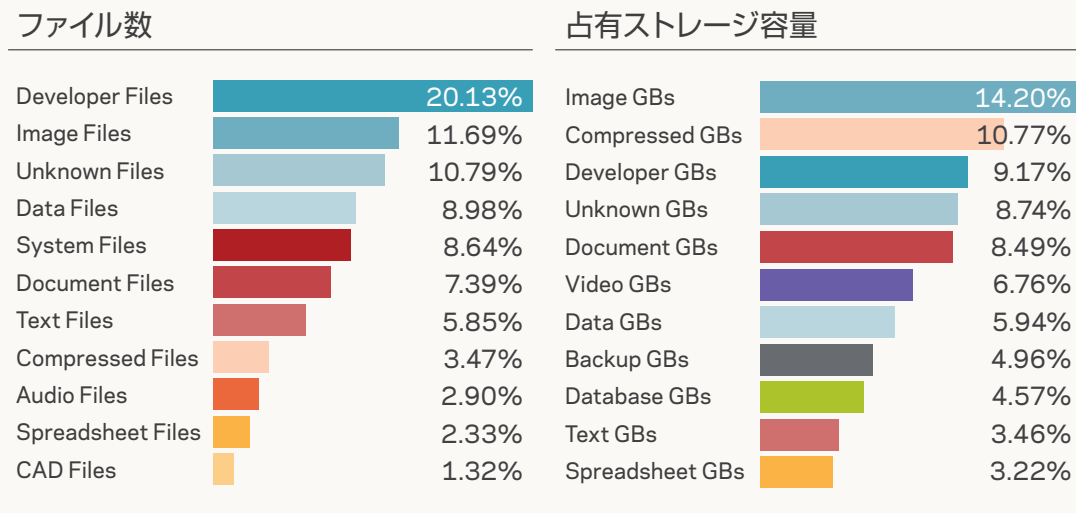
一方で必要となるストレージ容量は数の増加よりも 9% 高い割合で増え続けています。

データの伸びを抑えることは、単なるストレージ関連コストだけの問題ではありません。ストレージ環境にはあらゆるデータが野放図に散乱し、PB あたりの平均で 2,312,000,000 個のファイルが含まれているとされています。

利用者の善意に頼ることでデータの伸びをいくらかは抑えられるかもしれませんが、人海戦術での管理は限界です。抜本的なストレージ管理の改善が必要になってきています。

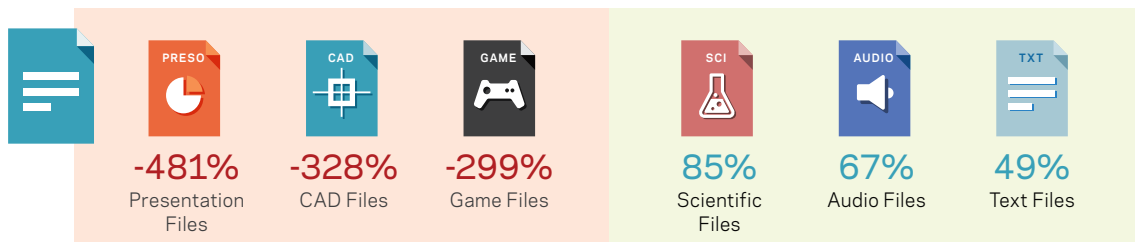
最大の障壁: 画像ファイルと開発中ファイル

すべてのファイルを分析した結果、数が最も多いファイルタイプと最も多くの容量を占有しているファイルタイプの間に違いがあることを発見しました。次の表はファイルのタイプ別に何がストレージ関連コストに影響しているかを示しています。

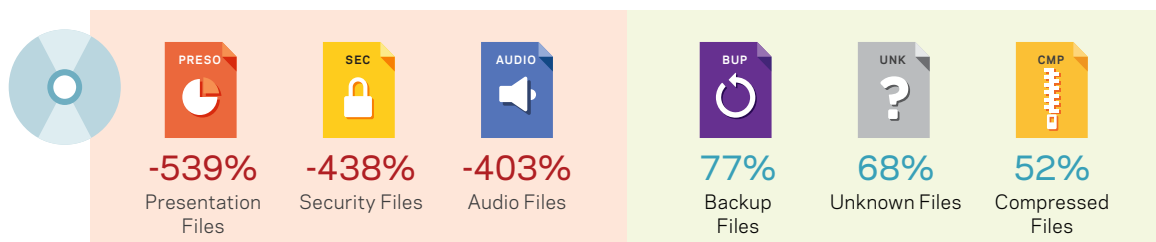


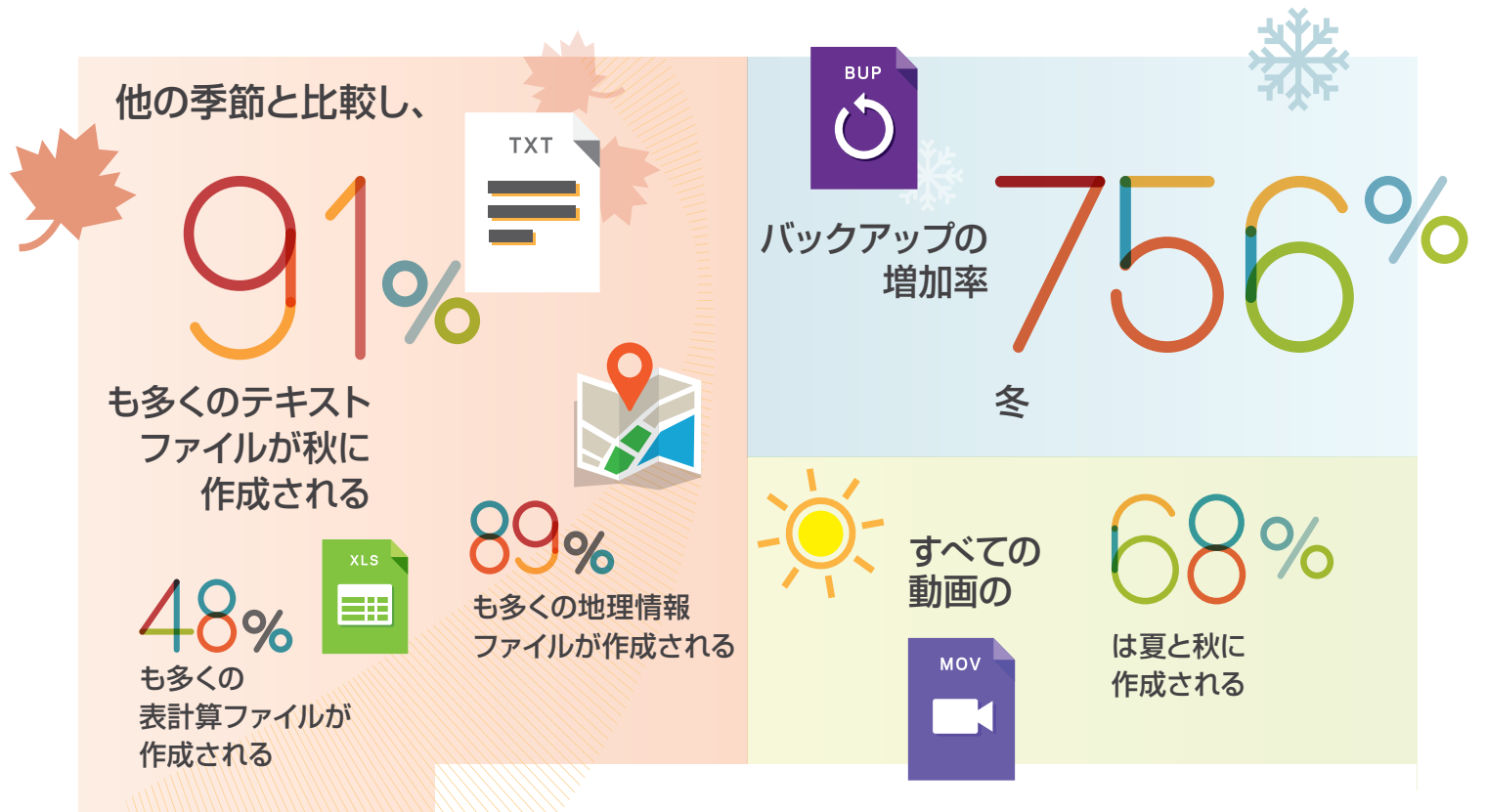
ファイルタイプの世代交代

上記のファイルタイプ別分析は時代とともに変遷します。過去10年間で他のファイルタイプと比較して数の変化が激しかったものは次のとおりです。



同様に占有ストレージ容量で見た場合、変化が激しかったものは次のとおりです。



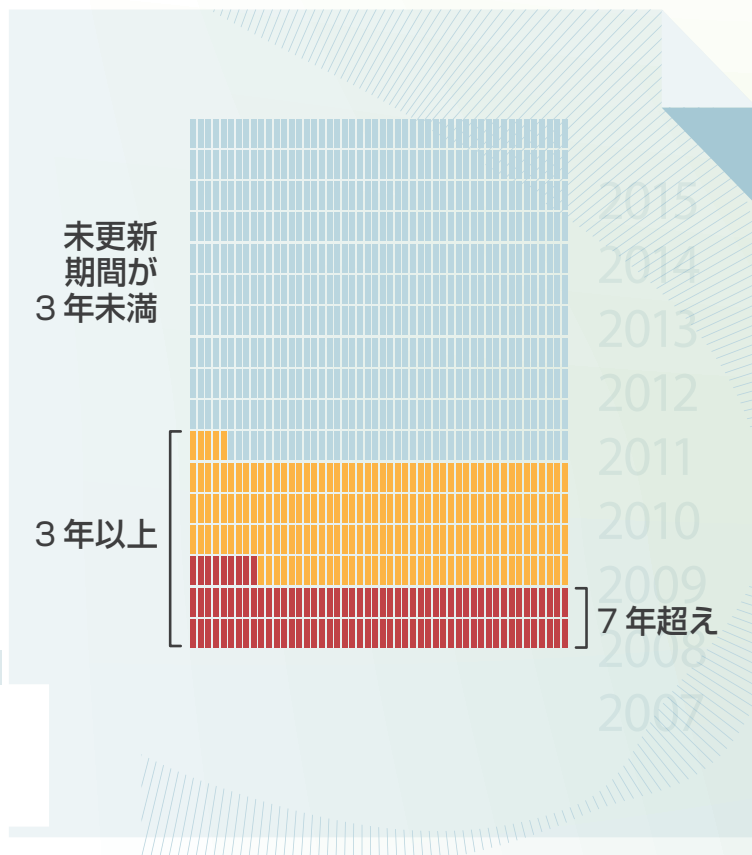


データ増加の季節変動

純粋なファイル作成という点では、秋に最も多くのファイルが作成されています。他の季節と比較し多くの企業の年度末に近い為、テキストファイルは 91% 増、表計算ファイル 48% 増、地理情報システムファイルは 89% も多くが秋に作成されています。

秋から冬で作成数が増えるファイルタイプは、バックアップと文書ファイルです。ファイルサイズで比較すると、バックアップは 756% の増加です。これは主に税務処理の証拠保全を目的とした財務部門からの法的要件による年次バックアップ処理が加わるためです。文書ファイルの多くも税務処理関連が増加します。

動画の 68% は夏と秋に撮影されており、冬には急激に 63% 減となります。電子メールは皆さんがいつも使っているので容易に予想されるとおりほぼ一定の増加で、季節間の増減に対する標準偏差はたったの 0.7% しかありません。



41%

のファイルは

3年間
更新されていない

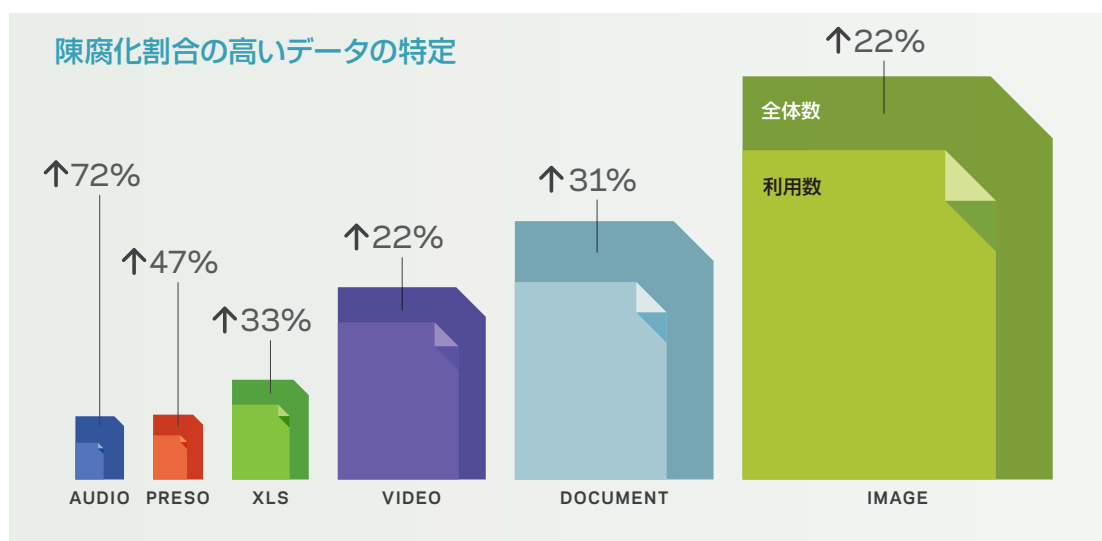
12% = 非常に古い
(7年間更新なし)

なぜ、すべてを保存しておくのですか？

現在のビジネスにとって情報ほど大事なものはありません。しかし、情報はあまりにも速いスピードで作成されているため、個々の情報の有用性が一過性のものになっています。

改善策: ファイルタイプの照準を定める

手に負えないほどの大量の陳腐化データを前に何らかの改善策の検討を強いられた場合は、コスト削減効果が高い順に優先順位を付けるとよいでしょう。全体数のうち陳腐化したデータの割合が突出しているファイルタイプを見ると、従来からある「office」製品の旧ファイルタイプが大きな負担をかけていることがわかります。



個人的なファイルの改善策をどこに適用すれば、ストレージ容量の面で最大のコスト削減効果を得られるのかを判断するには、1 ファイルあたりの容量削減効果が最も大きい、次の5つのファイルタイプに注目します。

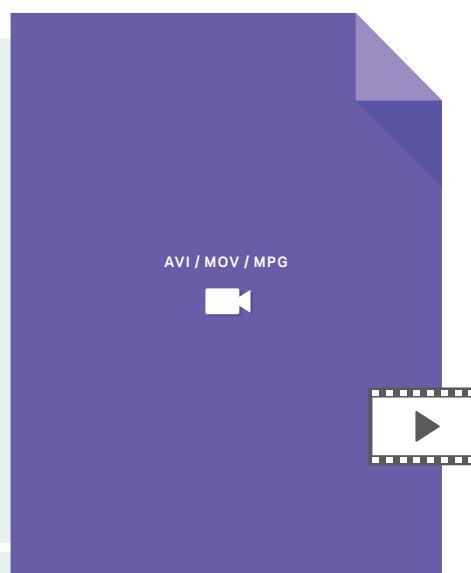
1. 仮想マシン関連のファイルタイプ
2. セキュリティ関連のファイルタイプ
3. ゲーム関連のファイルタイプ
4. 科学関連のファイルタイプ
5. 地理情報システム関連のファイルタイプ



動画ファイルは

15.8x

もの容量を消費する
(他の陳腐化ファイル
サイズに比べて)

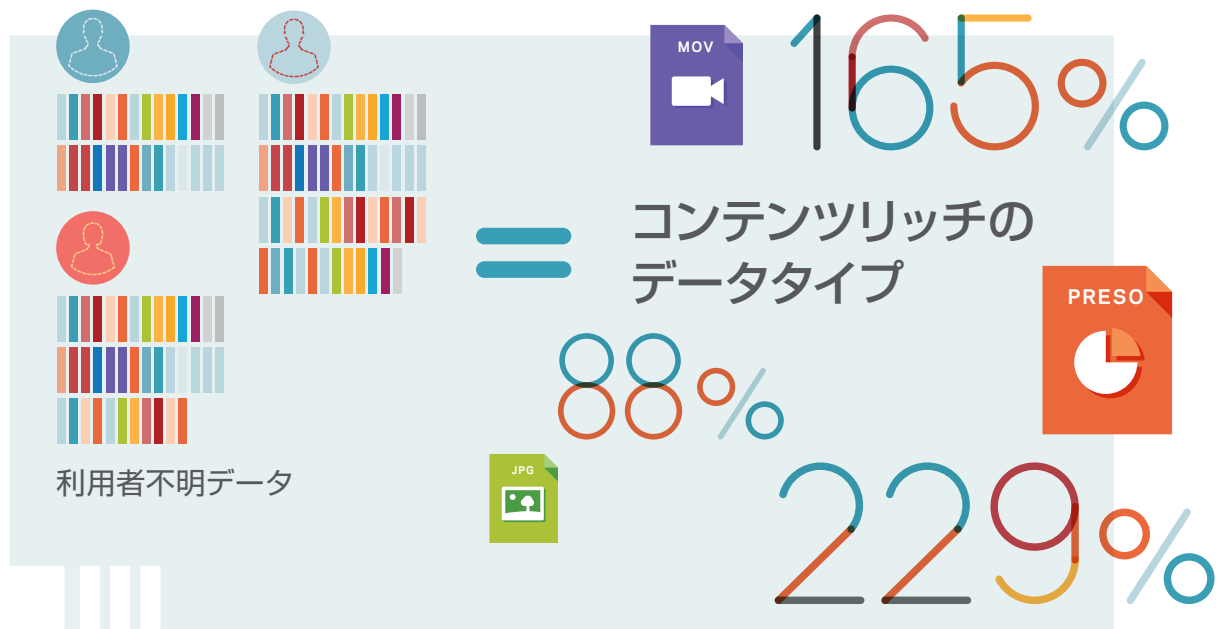


7.3x 6.4x 2.2x



不釣り合いな容量のファイルを特定する

ファイルタイプ別の改善策を検討するために、優先順位付けを行う場合は、ファイル数とファイル容量の釣り合いが取れていないものを探します。たとえば陳腐化した動画ファイルは、他の陳腐化ファイルと比べて 15.8 倍も多くなっています。同様にファイル数に比べて容量が大きいファイルとして、仮想マシン関連のファイルは 7.3 倍、プレゼンテーションは 6.4 倍、電子メールは 2.2 倍多くの領域を占有しています。この結果から、どのファイルタイプを優先すべきかが判明します。



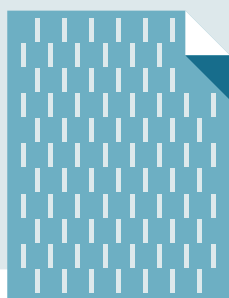
あの人、机の上だけは片付けたよ。 まあ、そんなところでしょう。

特定の所有者がないデータのことを所有者不明データと呼びます。従業員の入れ替わりや担当の変更、そしてよく見かける ActiveDirectory の無秩序状態などを考えれば、引き継がれなかった遺産を追跡することがどれだけ困難で、コスト面でも企業に大きな負担を強いることは容易に想像がつかます。

コストが生じる理由の 1 つは、所有者不明データが不釣り合いなストレージ容量を消費することにあります。全体のファイル数に占める所有者不明データの割合はたった 1.6% ですが、ストレージ容量全体に占める割合は 5.1% です。また、所有者不明データはコンテンツリッチのデータタイプの比重が高く、平均的なファイルに対して画像で 88%、動画は 165%、プレゼンテーションでは 229% も多く容量を消費しています。

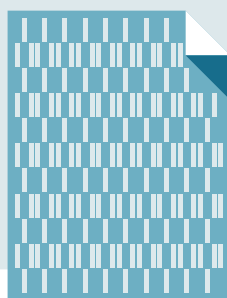
雇用の傾向もストレージ環境に影響しています。所有者不明データは所有者が特定できるデータと比べて容量が 222% 大きいという結果があります。ファイルのサイズが大きいほど重要なコンテンツが含まれていると考えた上司が、部下が去った後にサイズの大きなファイルだけを捨てずに保管しておくケースもあるようです。ストレージ容量の削減に向けた作業は、まずは所有者のいないコンテンツに的を絞って実施してください。

平均ファイル容量



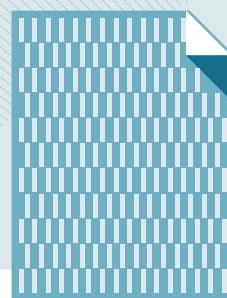
0.24 MB

過去 10 年間



0.40 MB

過去 5 年間



0.53 MB

過去 1 年間

密度は有用性の尺度にはならない

私たちが作成するデータの密度は日増しに高くなってはいますが、過去 7 年間の密度の伸びの速度はそれほど速くはなく、たった 10.3% しか増えていません。各ファイルの平均サイズは次のようになっています。

未使用期間が 10 年以上のファイル:	0.24MB
過去 5 年以内に使用されたファイル:	0.40MB
過去 1 年以内に編集されたファイル:	0.53MB

陳腐ファイルに分類されたファイルは過去 1 年間に編集されたファイルよりもサイズが **33% も小さい**という結果が出ています。

では、どうしますか？

もし、皆様の環境がこのレポートの調査対象になった環境に類似しているのであれば、ストレージコスト削減の大きなチャンスがあります。平均的な 10PB の環境を想像してみてください。

41% = \$20.5 million

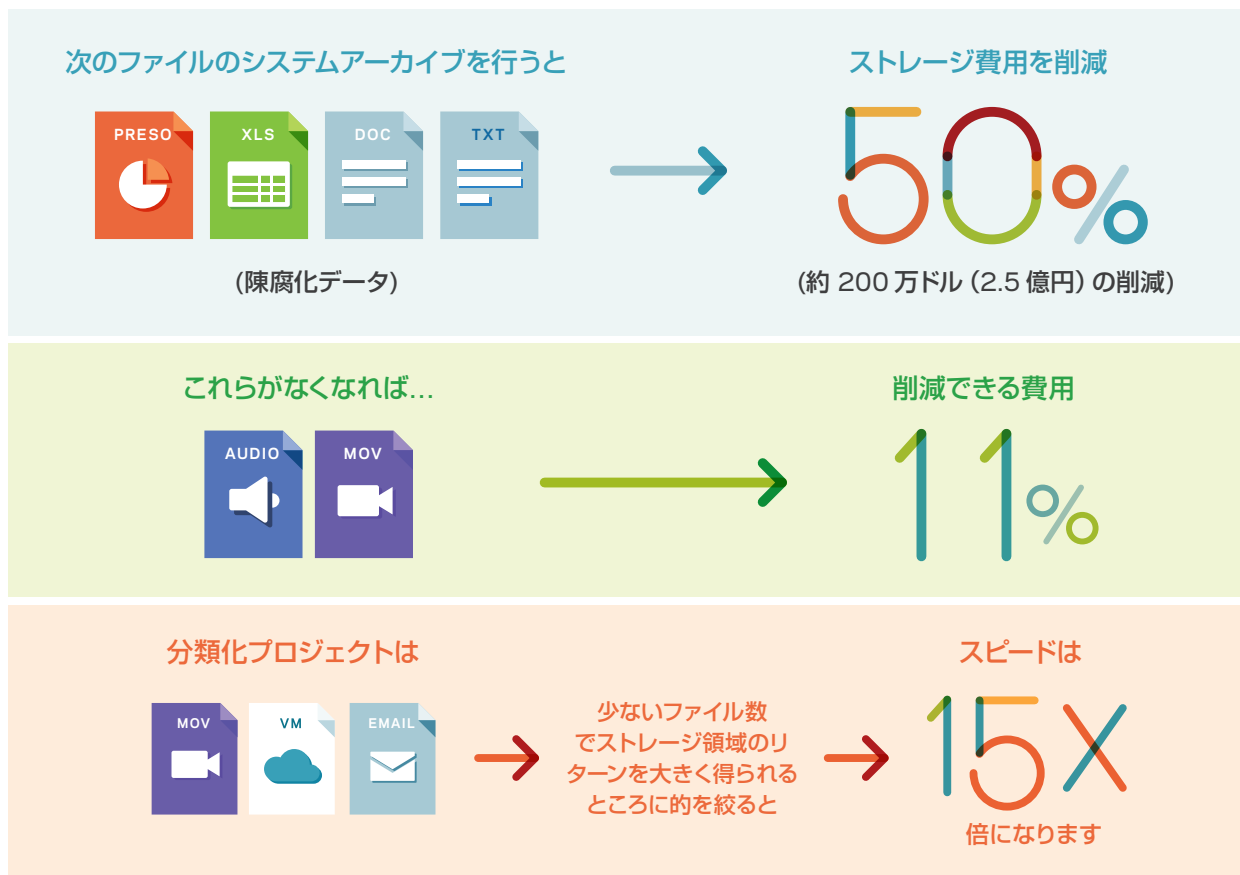
陳腐化データ 保管費用

このうちの 41% が陳腐化しているとしたら、過去 3 年間使用されていないデータに対し、年間 2050 万ドル (約 25 億円) も費やしていることとなります。とは言いつものの、整理することは大変な作業です。この 4.1PB という数字は、9,479,200,000 個のファイルについて分類、削除、アーカイブのいずれかの判断を強いられるということを示しています。

9,479,200,000

個別ファイル総数

優先順位を付けてください...



平均的な陳腐化環境の 20% は、プレゼンテーションや表計算ファイル、文書、テキストファイルなどコンテンツリッチなファイルが占めており、これらはファイルシステムのアーカイブ化プロジェクトの格好の対象です。50% 以上のストレージ費用の削減と 200 万ドル (約 2.5 億円) 以上の削減効果を見込めます。オーディオファイルと動画ファイルだけで、11% の削減効果があります。画像は 7 年以上前の古いカテゴリのストレージ容量の 18% を占めています。

少ないファイル数でストレージ領域の削減効果を大きく得られるファイルタイプ、たとえば動画や仮想マシン関連ファイル、電子メールに的を絞って、分類化プロジェクトを実施してください。そうすれば、15 倍速く始めることができます。従業員の退職や担当変更時のデータ保管ポリシーを評価することで 5%、つまり、100 万ドルもの削減効果が見込めます。

状況がどうであれ、これらの洞察によって凄まじいスピードで増えるデータに歯止めをかけ、健全な環境を取り戻すチャンスがたくさん生まれます。

VERITAS™

ベリタスは、世界最大規模の複雑な環境にも対応できるソリューションをご提供することで、お客様が情報の力を活用してビジネスを成功させるお手伝いをいたします。弊社のお客様の企業規模はさまざまで、Fortune 500社 に選出されたグローバル企業の 86% に製品をご採用いただいています。また、ベリタスは10年以上にわたり、ガートナーのマジック・クアドラントのエンタープライズバックアップソフトウェアおよび統合アプライアンス部門1とエンタープライズ情報アーカイブ部門2の両方で、リーダーとして認定されています。

データの可用性を高め、競争力を高める方法を見つけよう、ベリタスは経験豊富なパートナーコミュニティとともにお客様を支援します。世界 58 カ国に 7,800 人の従業員を擁するベリタスは、売上高 25億ドルの企業です。Amazon、Cisco Systems、富士通、Google、日立、HP、IBM、Microsoft、NetApp、OpenStackコミュニティ、Symantecなど、多数のテクノロジー大手企業とパートナー関係を築いています。

従来のデータセンターからプライベート、パブリック、ハイブリッドまで、ベリタスはパートナーコミュニティとともにインテリジェントな情報システムソリューションを提供して、どのような環境でもデータを適確に保護、識別、管理ができるようお客様を支援します。保有する情報を把握し、その保護方法を学び、削除すべき情報を理解するためのインサイト（洞察）と可用性をベリタスが提供いたします。

ベリタスについての詳細は、veritas.com でご確認ください。

利用規約:

出典: Gartner, Inc. 『Magic Quadrant for Enterprise Backup Software and Integrated Appliances (マジック・クアドラント エンタープライズバックアップソフトウェアおよび統合アプライアンス部門)』、Dave Russell、Pushan Rinnen、Robert Rhame、2015年6月15日。

出典: Gartner, Inc. 『Magic Quadrant for Enterprise Information Archiving (マジック・クアドラント エンタープライズ情報アーカイブ部門)』、Alan Dayley、Garth Landers、Anthony Kros、2015年10月29日。注: 2005年から2009年まで、本調査は「Magic Quadrant for Email Active Archiving (マジック・クアドラント 電子メールアクティブアーカイブ部門)」と呼ばれていました。

ガートナーは、ガートナー・リサーチの発行物に掲載された特定のベンダー、製品またはサービスを推奨するものではありません。また、最高の評価を得たベンダーのみを選択するようテクノロジーの利用者に助言するものではありません。ガートナー・リサーチの発行物は、ガートナー・リサーチの見解を表したものであり、事実を表現したものではありません。ガートナーは、明示または黙示を問わず、本リサーチの商品性や特定目的への適合性を含め、一切の保証を行うものではありません。

© 2016 Veritas Technologies LLC. All rights reserved. Veritas、Veritas ロゴは、Veritas Technologies LLC または関連会社の米国およびその他の国における商標または登録商標です。