

# ファイルサーバのデータ整理・活用を支援する「Information Assessment System」

室井 泰幸・向井 慶和

## 要 旨

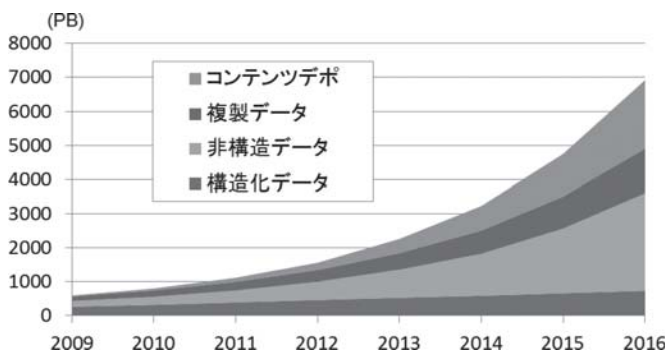
情報爆発が叫ばれるなか、ファイルサーバに格納される情報は肥大化の一途をたどっています。肥大化する情報により、ファイルサーバ内の情報把握と整理・活用は困難になってきています。ファイルサーバの「見える化」「スリム化」「活性化」「適正化」を実現するツールInformation Assessment Systemの最新版V2.1では、高速データ処理エンジンInfoFrame DataBoosterを活用することで、大規模ファイルサーバへの対応と高速検索・集計によるインタラクティブな分析を実現しました。

## キーワード

- ファイルサーバ ●肥大化 ●ファイル整理 ●情報活用 ●メモリ ●DB ●NAS
- 非構造化データ ●NIAS ●情報爆発

## 1. まえがき

企業内で活用されるデータは、データベースなどに格納される構造化データと、ファイルサーバなどに格納される文書・画像・ログなどの非構造化データに大きく分けられます。データ管理方法が規定されている構造化データに比べ、非構造化データは、標準的な管理方法が規定されていないため、データの増加に伴い情報管理が困難になる傾向があります。IDC Japanの調査発表<sup>1)</sup>によると、2011-2016年のストレージに格納されるデータ量の年間平均成長率は、構造化データの13.4%に対して、非構造化データは52.2%と非構造化データを中心にデータ量の



出典: IDC Japan「国内ディスクストレージシステム市場 2011年の分析と2012年~2016年の予測」  
TABLE112:国内ディスクストレージシステム・コンサンクションモデル セグメント別出荷容量予測、2009年~2016年

図1 データ種別ごとのストレージ総容量推移

爆発的な増加が見込まれています（図1）。

非構造化データが爆発的に増加するのに伴い、ファイルサーバの管理を人手で行い続けることは、もはや不可能に近く、肥大化し続ける情報の整理・活用はますます困難になっています。

本稿では、ビッグデータの中でも各企業に必ず存在し、かつ容量増加が顕著な非構造化データに着目し、ファイルサーバの状況を把握・分析、格納データの整理・利活用を容易にするツールとして開発した「Information Assessment System」のコンセプトと機能紹介を行います。

## 2. 開発の背景とファイルサーバが抱える課題

NECでは、お客様のファイルサーバ状態を確認して利用状況を分析し、利用状況に応じた適切な情報管理やストレージ増設方法を提案する「情報アセスメントサービス」を、これまで100社以上に提供してきました。

この情報アセスメントサービスをご利用いただいたお客様から、単なるストレージ増設だけではファイルサーバに蓄積されたデータの利活用が進まないため、アセスメントで分析した情報を基にしたデータの有効活用や、不要ファイルの削除・整理をしたいという要望を多数受けました。こうした要望に応えるため、お客様ご自身で使用して情報活用の促進に役立てていただくツールの開発を進めました。

### ファイルサーバが抱える課題

前述のとおり、爆発的に増大する非構造データを格納するファイルサーバに対して、運用管理者観点、利用者観点で次に示す課題が顕在化しています。

運用管理者観点での大きな課題は、ファイルサーバの状況把握（見える化）が困難になっていることです。ストレージの物理容量の拡大と爆発的なデータ増加が相まって、ファイルサーバの的確な利用状況把握が難しくなっています。しかし、利用状況を的確に把握していないと、次のようなケースに即応することができなくなります。例えば、ファイルサーバの空き容量が切迫した場合、即座にファイルサーバを増設することは難しいため、ファイルを削除・移動してスリム化し、空き容量を確保する対応が一般的です。しかし、利用状況の把握が十分でないことや、運用管理者だけでは格納情報の要否を判断することが難しいため、ファイルの削除・移動は容易ではありません。更に、利用者に確認を取るにも、膨大なファイルの中から整理するファイルを選別する必要があり、作業に手間を要します。

また、情報セキュリティ確保の観点から、ファイルサーバのアクセス権の適正管理（健全化）が重要になりますが、すべてのアクセス権の設定状況を把握し、健全な状態を維持するのは容易ではありません。企業内のファイルサーバ運用において、情報セキュリティを確保するためには、組織・プロジェクト単位でアクセス権を適切に設定する必要があります。しかし、エクスプローラの権限管理機能では権限状態を一括確認できないため、膨大なファイル・フォルダ権限の設定状況を個々に確認していく必要があり、多くの手間が掛かります。

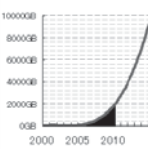
ファイルサーバ利用者観点での課題は、情報の活用（活性化）が困難になっている点が挙げられます。あるはずのファイルが見つからない、同じようなファイルが複数あってどれが最新か分からないなど、肥大化する情報のなか、格納された情報の検索性は年々低下し、情報の見通しが利かなくなっています。玉石混交状態のファイルサーバから重要かつ有益な情報をいかに探し出し利活用するか、すべての利用者の大きな課題となっています。

### 3. Information Assessment Systemの提供機能


ファイルサーバに関する課題に対して、お客様の要望に応

#### ①見える化

**増加傾向**  
例) 5年後は9TBに増加




**参照傾向**  
例) 78%は90日以上未参照のファイル



✓ **ファイルサーバの利用状況をグラフやレポートで見える化できます。**  
・増加、参照、更新、重複、拡張子、所有者の傾向を指定単位で自由にレポートング

#### ②スリム化


**不要ファイルリスト**  
不要ファイルリストを出力し画面上から削除・移動・圧縮の対応が可能。




✓ **不要（未参照・重複）ファイルを一括して削除・移動・圧縮を実現。**  
・管理者からユーザーへ整理依頼あるいはポリシー設定による自動整理が可能

#### ③活性化

**ファイル高速検索**  
ファイル名、プロパティ情報をキーにしてサーバ内のデータを高速に検索




**類似ファイル検索**  
類似ファイルを提示してデータ再利用を活性化



✓ **ファイルサーバに負荷をかけない検索・類似検索によりサーバの情報活用を促進**

#### ④適正化

**アクセス権限の設定状況を可視化し、適切に権限を割り当て、ユーザー・ファイル単位でのアクセス権を健全化**



✓ **アクセス権可視化**  
フォルダ・ファイルのアクセス権設定状況を可視化

図2 Information Assessment Systemの特長

えるため、前述の情報アセスメントサービスでも実施していた、ファイルサーバの情報把握を行う「見える化」に加えて、肥大化した情報を整理するための「スリム化」、ファイルサーバに格納された情報の利用促進を行う「活性化」の3つの機能を柱とした、ファイルサーバ整理・活用ツール「Information Assessment System V1.1」を2011年8月に製品化しました。

2012年7月に出荷されたInformation Assessment System V2.1では、上記に加え、ファイルサーバの権限設定状況の可視化と再設定を行うファイルサーバの「適正化」機能が追加されています（図2）。

### 3.1 見える化

ファイルサーバの整理・活用を行うためには、まずはファイルサーバの利用状況の把握が重要になります。Information Assessment Systemでは、ファイルサーバからファイルの格納情報

を収集して、利用状況をグラフやレポートで「見える化」します。見える化により、ファイルサーバの全体状況を把握するとともに、さまざまな角度からファイルサーバの肥大化要因を確認できます。例えば、少数のファイルがディスクの大部分を占めている、一定期間参照/更新のないファイルや重複文書が多数存在するといった傾向を読み取り、対策検討の参考とすることができます。

### 3.2 スリム化

ファイル整理は、ディスク容量削減のみならず、不要な情報が整理され情報の見通しを良くすることで、情報の検索性を向上させる効果があります。ファイルサーバ整理による「スリム化」を行うため、Information Assessment Systemでは、不要ファイルを特定の条件で絞り込み、削除・移動・圧縮によりファイルサーバを整理する機能を提供しています。膨大なファイルから不要ファイルをすべて確認することは、現実的ではありません。より少ない作業で効果的にディスク容量を削減するために、整理条件を調整して、整理対象文書を現実的に確認可能な分量にまで絞り込み、策定した整理指針を展開することで、効率的にファイルを整理できます。

整理に当たって、運用管理者が個々のファイルの必要性・重要性を確認することは困難なため、Information Assessment Systemでは、3階層の管理権限での整理に対応しています。ファイルサーバの運用管理を行う「システム管理者」、部門単位で管理を行う「グループ管理者」、「一般利用者」の3階層で整理依頼と確認を行い、階層間で相互連携したファイル整理を実現します。

また、手動整理に加え、ファイルの自動整理にも対応しています。あらかじめ整理条件を設定しておくことで、整理条件に適合したファイルを定期的に自動整理します。自動整理機能を利用することで、ポリシーに従い一定期間アクセスされていないファイルを移動する、一時領域フォルダを定期的に削除するといった自動運用が可能になります。

更に、ファイル移動による整理では、iStorage HSなどの二次ストレージに移動させることで、ストレージの持つ重複排除・物理圧縮機能を有効活用し、物理容量を削減するとともに、効果的な情報格納を支援します。

### 3.3 活性化

ファイルサーバに格納されている情報を有効活用するには、

膨大な格納情報の中から、欲しい情報を探して見つけ出す必要があります。こうした情報検索ニーズに対応するため、ファイル名や更新日付など、さまざまな条件を指定して目的のファイルを探し出す機能を提供します。ファイルサーバ利用者がこの機能を利用することで、日常業務でのファイルサーバの利活用を促進します。

### 3.4 権限設定の適正化

ファイルサーバのアクセス権の設定状況を可視化し、不適切なアクセス権を見つけ出し再設定することで、ファイルサーバの権限状態を適正化する機能を提供します。

ファイル・フォルダ階層間で、不適切に設定されたアクセス権や、特定ユーザーがアクセスできるファイルを一括確認でき、ファイルサーバの所有権変更機能などと合わせて、人事異動時などのアクセス権限再設定に対応できます。

## 4. 情報爆発への対応

### 4.1 従来バージョンでのデータ管理の課題

Information Assessment System V1.1（以下、V1.1）では、ファイルサーバの情報管理にリレーショナルデータベース（以下、RDB）を使用していました。ファイルサーバから収集した膨大なデータ処理に多くの時間が掛かるため、他社製品と同様に、情報収集時に決められた条件でバッチ処理によるデータ集計を行う必要がありました。情報爆発が進むなか、数十TBの容量を持つファイルサーバも珍しくなくなってきましたが、V1.1では、RDBの性能限界により1台の管理サーバで対応できるファイルサーバの最大容量の目安を10TB程度としていました。10TBを超えるファイルサーバには、管理サーバの複数台構成で対応していましたが、管理用サーバ自身の導入運用コストが増大するという課題がありました。

### 4.2 リアルタイム集計の実現

Information Assessment System V2.1（以下、V2.1）では、データ管理基盤を見直し、メモリデータベースを用いてRDBが不得意とする大規模データを一括高速処理できる、高速データ処理エンジン「InfoFrame DataBooster」を採用しました。

これにより、データの集計速度をV1.1比で約40倍と飛躍的に高速化し、V1.1では処理速度の問題から対応できなかった、インタラクティブなファイルサーバの分析・集計を実現しています。ファイル整理条件をその場で細かく変更して整理対象を絞り込むことで、効果的なファイル整理ポリシー設定が可能になりました。

#### 4.3 大容量データへの挑戦

また、InfoFrame DataBoosterの採用により、1台の情報管理サーバで取り扱い可能な最大容量の目安を、V1.1の10TBから、V2.1では50TBに拡張し、大規模ファイルサーバの高速分析・集計に対応しました。50TBのデータは、1ファイル当たりのデータ量を500KBと仮定すると、1億ファイルに相当します。これらの膨大なデータを、Windows Server 2008 R2 Standard Editionのメモリ上限となる32GB以内で処理することを目標に設計しています。

ファイルサーバ検査と情報集計に必要なすべての管理データをInfoFrame DataBoosterに格納すると、ソフトウェアの動作メモリ量を増加させることになり、このメモリ要件が非現実的な値となってしまいます。そのため、ファイルサーバの情報集計・情報検索に必要な情報をInfoFrame DataBoosterに格納し、それ以外の情報をRDBやファイルで併用管理するハイブリッド情報管理方式を採用しました。

更に、InfoFrame DataBoosterの重複データ圧縮機能を生かし、ファイルの権限情報など、重複するケースが多く想定されるデータをメモリデータベース上で管理することで、集計検索の高速化とメモリ使用量削減を実現しています。

また、性能面の配慮から、情報の追加・更新・削除を行う際は、複数のデータ処理をひとまとめにし、一括処理することで処理性能の向上を図っています。

### 5. むすび

情報爆発のなか、ファイルサーバに格納されている情報は肥大化を続けています。メモリデータベースを活用して大規模データに対応した製品技術と、情報アセスメントサービスで培ったノウハウを生かしたInformation Assessment Systemを軸に、企業内の情報管理と情報活用のあり方を、今後も提案していきます。

\*Windows、Windows Serverは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。

#### 参考文献

- 1) IDC Japan : “国内ディスクストレージシステム市場 2011年の分析と2012年～2016年の予測,” J12430103, 2012.5

#### 執筆者プロフィール

室井 泰幸  
ITソフトウェア事業本部  
第三ITソフトウェア事業部  
マネージャー

向井 慶和  
ITソフトウェア事業本部  
第三ITソフトウェア事業部  
主任

#### 関連URL

Information Assessment System製品情報:  
<http://www.nec.co.jp/soft/ias/>

# NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。  
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報(日本語)

NEC Technical Journal(英語)

## Vol.65 No.2 ビッグデータ活用を支える 基盤技術・ソリューション特集

ビッグデータ活用を支える基盤技術・ソリューション特集よせて  
ビッグデータを価値に変えるNECのITインフラ

### ◇ 特集論文

#### データ管理/処理基盤

超高速データ分析プラットフォーム [InfoFrame DWH Appliance]  
SDN 技術で通信フローを制御する [UNIVERGE PF シリーズ]  
大量データをリアルタイムに処理する [InfoFrame Table Access Method]  
大量データを高速に処理する [InfoFrame DataBooster]  
ビッグデータの活用最適なスケールアウト型新データベース [InfoFrame Relational Store]  
高い信頼性と拡張性を実現した Express5800/ スケーラブル HA サーバ  
大規模データ処理に対する OSS Hadoop の活用  
大容量・高信頼グリッドストレージ iStorage HS シリーズ (HYDRAStor)

#### データ分析基盤

ファイルサーバのデータ整理・活用を支援する [Information Assessment System]  
超大規模バイオメトリック認証システムとその実現  
WebSAMの分析技術と応用例～インバリエント分析の特長と適用領域～

#### データ収集基盤

スマートな社会を実現する M2M とビッグデータ  
微小な振動を検知する超高感度振動センサ技術開発とその応用

#### ビッグデータ処理を支える先進技術

多次元範囲検索を可能とするキーバリューストア [MD-HBase]  
高倍率・高精細を実現する事例ベースの学習型超解像方式  
ビッグデータ活用のためのテキスト分析技術  
ビッグデータ時代の最先端データマイニング  
ジオタグ付きデータをクラウドでスケラブルに処理するジオフェンシングシステム  
柔軟性と高性能を備えたビッグデータ・ストリーム分析プラットフォーム [Blockmon] とその使用事例

### ◇ 普通論文

地デジ TV を活用した「まちづくりコミュニティ形成支援システム」

### ◇ NEC Information

#### NEWS

スケールアウト型新データベース [InfoFrame Relational Store] が 2 つの賞を受賞



Vol.65 No.2  
(2012年9月)

特集TOP