

# 1000万件のメールを1秒で検索する “LogAuditor Mail Saver”

大塚哲史\*  
石川雅朗\*  
加藤 守\*\*

“LogAuditor Mail Saver” : Email Archive Solution with High-Speed Search up to 10 Million Emails per Second  
Tetsufumi Otsuka, Masaaki Ishikawa, Mamoru Kato

## 要 旨

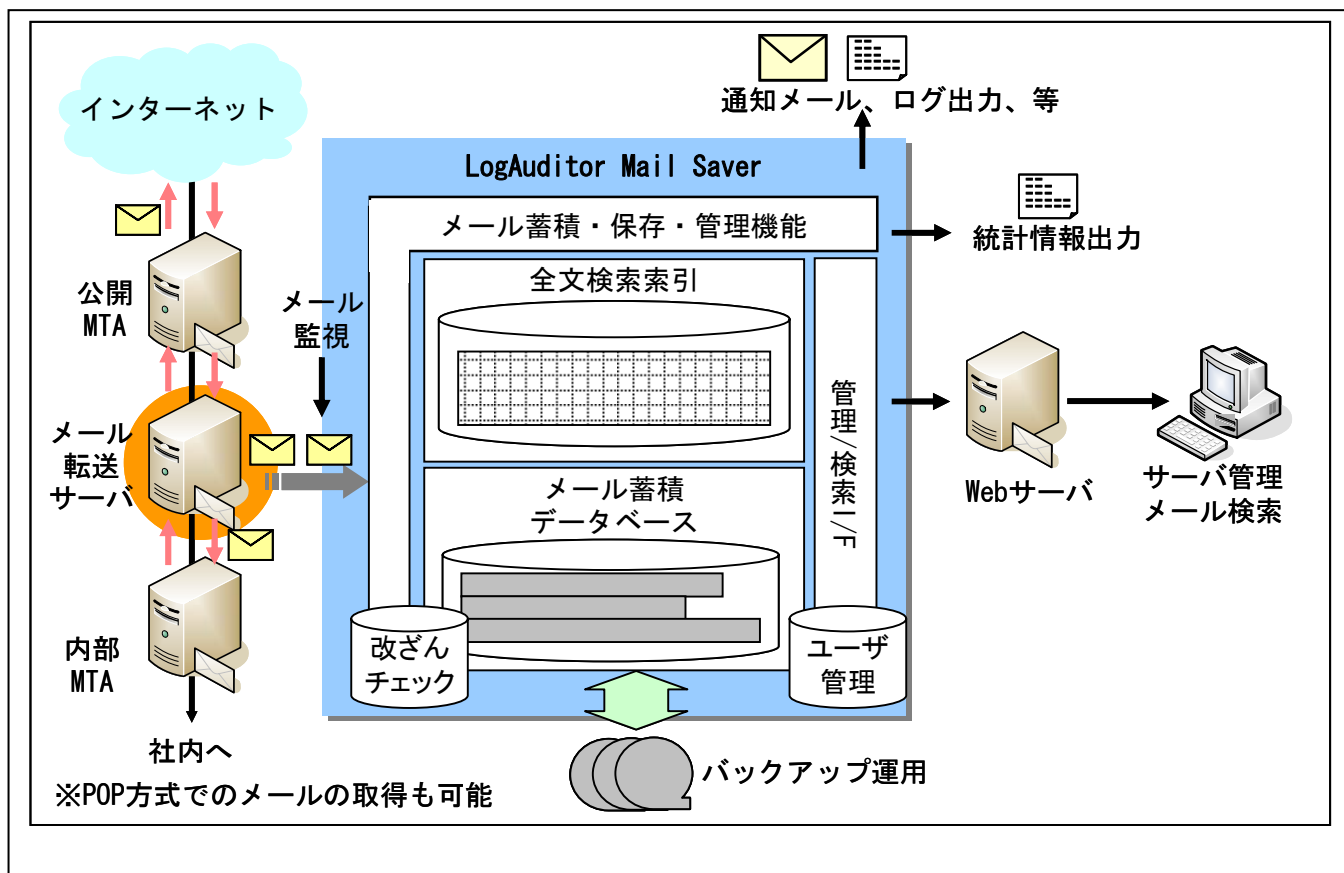
ビジネスツールとして日々利用されている電子メールには、企業の様々な情報をはじめ、重要な情報が数多く含まれている。メールは容易に利用できる反面、情報漏洩などの要因にもなり得る。メールの運用監視策としてメールを長期間保管し、必要な時に監査できるメールアーカイブシステムが必須となっている。メールアーカイブシステムの課題としては、大量のメールを保管する為のストレージコスト、大量のメールから目的のメールを取り出す為の検索速度、利用者の権限の範囲内で参照するセキュリティの確保などの点にあった。

三菱電機インフォメーションテクノロジー(株)(MDIT)が提供するメールアーカイブソリューション “LogAuditor<sup>(注1)</sup>”

Mail Saver” は、三菱電機独自の技術により、大量のメールの高速蓄積、データ圧縮によるストレージ容量削減、高速全文検索を実現した製品であり、以下の特長がある。

- (1) 1TBのメール（平均メールサイズ約100KBのメールで約1000万通に相当）の中から目的のメールを1秒で検索可能な高速全文検索
- (2) 全文検索索引を含めたアーカイブデータを元メールから最大20%圧縮してストレージ容量を削減
- (3) ユーザ/ロール/メールセットビューの組み合わせにより、利用者に与えられた権限の範囲でメールの検索/表示が可能なユーザ管理を実現

(注1) LogAuditorは三菱電機インフォメーションテクノロジー(株)の登録商標である。



## “LogAuditor Mail Saver” のシステム構成

LogAuditor Mail Saver は大量のメールを蓄積・保存し管理する機能、大量の非定型データの圧縮・蓄積機能を提供するメール蓄積データベース、高速全文検索エンジンの全文検索索引から構成され、管理/検索用フロントエンドとして Web ブラウザで動作する運用管理画面、検索画面を提供する。

## 1. ま え が き

ビジネスツールとして日々利用されている電子メールには、企業の様々な情報をはじめ、重要な情報が数多く含まれている。メールは容易に利用できる反面、情報漏洩などの要因にもなり得る。内部統制、リスク管理、コンプライアンスの観点から、メールを長期間保管し、必要に応じて迅速にメールを取り出せるメールアーカイブシステムが必要となってきた。

メールアーカイブシステムとは、送受信した全てのメールを1箇所にまとめて保存するシステムを指す。通常は保存してから一定期間は効率的な検索が可能なハードディスクに保存し、一定期間後に順次バックアップメディアへの保存を行うという運用がとられる。

本稿では、従来のメールアーカイブシステムにおける課題、三菱電機インフォメーションテクノロジー(株)で製品化したメールアーカイブシステムLogAuditor Mail Saverの特長、及び、製品を支えるデータ高速処理技術について述べる。

## 2. メールアーカイブシステムの課題

メールの容量は添付ファイルなどを伴って増加傾向にあり、送信されるメール数は1998年から2006年までの間に3倍に増加しているという調査結果<sup>(1)</sup>もある。最近では、大企業で送受信されるメールは、年間で数十テラバイト規模に及ぶ事例も見られる。大量のメールを取り扱う必要があるため、従来のメールアーカイブシステムでは以下のような課題があった。

- ・ 監査や追跡調査に必要なメールの検索に長時間を要し、効率的な追跡調査が困難
- ・ 長期保管に必要なストレージコストが高い
- ・ 全てのメールを一括保管するため、適切な権限によりメールの参照範囲を制限する必要がある

LogAuditor Mail Saverはこれらの課題を解決し、大量のメールの長期保管と迅速な検索を可能とする。

## 3. LogAuditor Mail Saverとは

### 3.1 LogAuditor Mail Saverの特長

LogAuditor Mail Saverは、大量のメールの蓄積・保存・検索を可能とするための機能を提供する製品である。蓄積するメールはメール転送サーバから取得するSMTP方式とジャーナル機能で蓄積されたメールを取得するPOP方式でのメールの取得が可能である。主に、以下のような特長がある。

#### (1) 大量メールデータの高速全文検索

1TBのメール(平均メールサイズ約100KBのメール約1000万件)のヘッダ・本文・添付ファイルの内容を対象とし

て、約1秒で検索する高速全文検索エンジンを搭載し、長期保存している大量メールへの高速検索を可能としている。

#### (2) メールストレージコスト削減

従来のメールアーカイブシステムでは、全文検索索引を含めるとストレージ容量が元メールサイズの1.5倍程度に増大するものもあり、大量のメールを長期保存するには、数十テラバイトにも及ぶ大規模なストレージシステムが必要であった。LogAuditor Mail Saverは、全文検索索引の効率化、データ圧縮技術により、全文検索索引を含め、元メールより約20%容量を削減でき、大量メールの長期保存に対するストレージ容量の大幅な削減を可能としている。

#### (3) ユーザ管理機能

メールアーカイブシステムへのアクセス管理のみならず、昨今の内部統制、セキュリティ管理要件に対応するため、ユーザ/ロール/メールセットビューの組み合わせにより、利用者に与えられた権限の範囲でメールの検索/表示が可能なユーザ管理を実現している。これにより、上長が部下の送信メールのみを監査する運用を可能とし、人事異動、組織変更などに伴うユーザ管理の変更に対応可能である。

図1に、2つの課から構成される営業部において、部下のメールのみを監査する権限の設定例を示す。営業一課ビューとして、営業一課員から送信されたメールのみを検索対象とする設定を行い、営業一課ビューを使用するロールとして営業一課長ロールを設定する。ロールの使用者として営業一課長自身を設定することで、営業一課長は営業一課の課員のみメールの監査権限が設定される。

同様に、営業二課ビューを設定し、営業一課、営業二課ビューを営業部長ロールとして割り当てることで、営業部長が営業部員のみを対象としたメール監査権限が設定される。

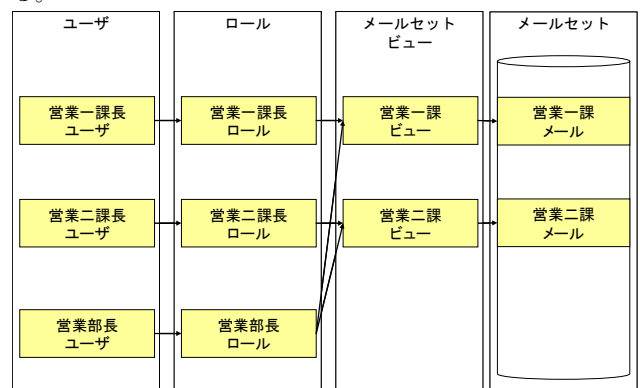


図1. ユーザ管理

### 3.2 LogAuditor Mail Saverの主要機能

#### (1) 漏れのない全文検索機能

LogAuditor Mail Saverの全文検索機能は、カタカナの拗音/促音、ひらがなの拗音/促音、英字の大文字/小文

字、文字の全角/半角などの異表記を同一視した検索や、検索キーワード内の途中の空白や改行を無視した検索にも対応し、漏れのない検索を実現している（図2）。

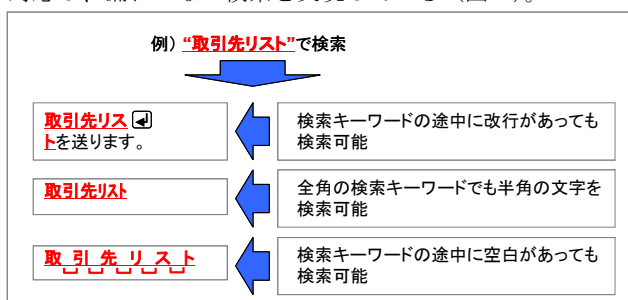


図2. 異表記検索

(2) メール監視機能

メール内容の監視機能により、メールのアーカイブ時に予め設定した監視条件に合致するメールが送信された事を検知し、ログ出力や管理者への通知等を行うことが可能である。これにより、情報漏洩などの事故発生時に迅速かつ効率的に対応が可能となる。

(3) メールの証拠能力を保証する改ざんチェック機能

メール蓄積データベースは、追記型データベースで構成され、上書き、更新ができないシステムとなっている。また、アーカイブ時にハッシュ値を取得し、後日現在のメールデータとハッシュ値を照合することにより、改ざんの有無を検出し、アーカイブメールの真正性を保証することが可能である。

(4) バックアップ機能

メールを月/日単位などの時系列的な“範囲”に分割して管理し、“範囲”を単位としたバックアップ/リストア、削除が可能である。オンラインでの保存期間を過ぎた“範囲”をテープなどにバックアップして削除し、必要に応じてリストアできるようにすることで、ストレージの効率的利用が可能となる（図3）。

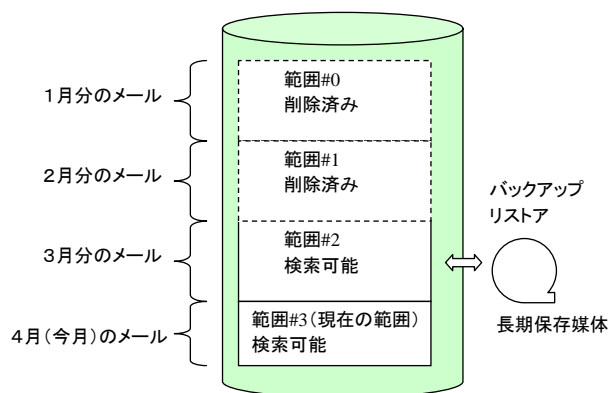


図3. 範囲の概念

(5) 統計情報出力機能

LogAuditor Basic/Enterprise<sup>(2)</sup>と連携することで、メールアドレス別送受信件数、日付別送受信件数、時間帯

送受信件数

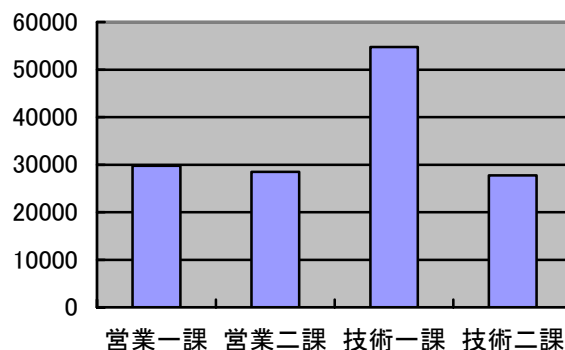


図4. レポート出力例

別送受信件数などをレポート出力することが可能である（図4）。

(6) Webブラウザによる管理/検索

LogAuditor Mail Saverの管理/検索については、Webブラウザによる画面操作で完結する形で設計されており、クライアント側は、Webブラウザ以外は不要となっている（図5）。

メールセットビュー	営業一課 営業二課 営業三課
期間(送信日時)	2007年10月1日から 2007年11月11日まで
並び替え(送信日時)	●指定無し ○昇順 ○降順
送信アドレス(From)	
受信アドレス(To, CC)	
件名(Subject)	
添付ファイル名	
添付ファイル有無	●指定無し ○有り ○無し
キーワード (本文、添付テキスト)	
最大件数	1000件

図5. 検索画面イメージ

また、管理/検索機能インターフェースとしてJava<sup>(注2)</sup> APIを提供しており、必要に応じてユーザインターフェースの変更、他のシステムとの連携等に柔軟に対応が可能である。

4. LogAuditor Mail Saverの高速処理技術

LogAuditor Mail Saverは当社独自の大規模データ高速処理アーキテクチャSISA (Scalable Intelligent Storage Architecture)に基づき、大量のメールの高速処理と効率的な管理を実現した。以下にその主要技術について示す。

4.1 構成

メールアーカイブには以下のような特長がある。

- ・ サイズが可変長であり、従来のデータベースでは効率的に管理できない

(注2) Javaは、米国Sun Microsystems, Inc.の登録商標である。

- ・ メール件数が多く、通常のファイル管理ではオーバーヘッドが大きい
- ・ 大量に蓄積されるため、圧縮によるストレージ容量削減が必要
- ・ 時系列に追加されるため、更新は不要
- ・ 日付、件名、アドレス等の属性による検索やキーワード検索機能が必要

このようなメールアーカイブの特長に対応するために、ログデータベースLDB(Log DataBase)<sup>(2)</sup>と高速全文検索エンジンFTS(Full Text Search)<sup>(3)</sup>という2つの技術を組み合わせた当社独自のメールデータ管理用データベースを構築した。図6に構成を示す。

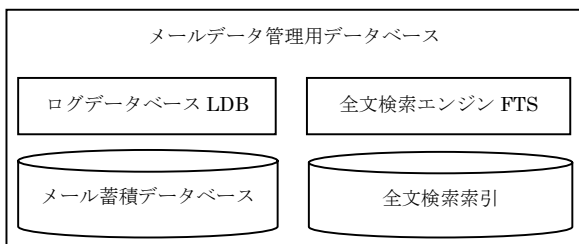


図6. メール管理用データベースの構成

ログデータベースLDBは、時系列的に追加される可変サイズの非定型データを効率的に管理可能な追記型データベースであり、メールを保管するメール蓄積データベースを管理する。高速全文検索エンジンFTSは、キーワード検索を可能とする全文検索索引の管理を行う。

蓄積時にはメールをメール蓄積データベースに自動的に圧縮して保管すると同時に、メールからテキストと属性情報を抽出して全文検索索引を生成する。検索時には全文検索索引によりメールIDを取得した後、そのメールIDによりメール蓄積データベースからメールを取得することが可能である。

#### 4.2 高速全文検索

全文検索エンジンFTSはブロック化n-gram索引方式により高速化を実現している。ブロック化n-gram索引とは、広く利用されているn-gram索引のデータ配置に独自の工夫を加えたものであり、以下の特長を備えている。

- ・ 最小限のデータ読み出し
- ・ I/O単位の最適化（バッファサイズの範囲で最大のI/O実行）
- ・ 一方向のデータ読み出し（シークを最小化）
- ・ 複数ストレージへの分散

全文検索索引、メール蓄積データベースともに、複数ストレージへのデータ分散配置と並列処理により高速化が図られている。全文検索索引では索引の読み出し処理や照合処理を並列に実行し、メール蓄積データベースでは読み出し処理/圧縮・伸張処理を並列化する。これにより、プロセッサ数やストレージ数に応じて処理速度を向上させるこ

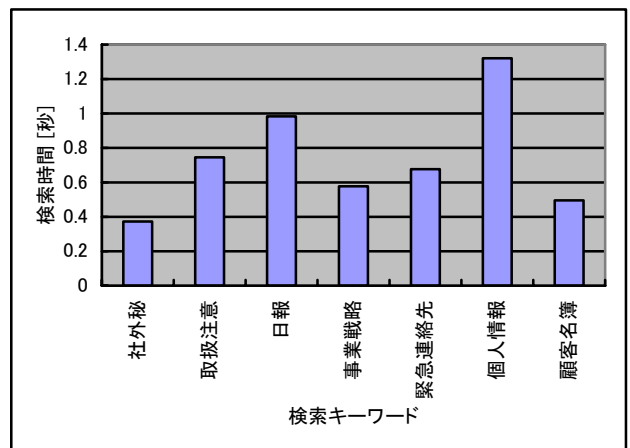


図7. メール1TBの検索時間

とができ、データ規模に応じた処理速度を実現している。

図7に検索時間の一例を示す。1TBのメール（平均メールサイズ約100KBのメール約1000万件）を約1秒で検索することが可能である。

## 5. むすび

メールを長期間保存し、迅速に取り出す事ができるメールアーカイブソリューションLogAuditor Mail Saverについて紹介した。今後、内部統制やリスク管理の強化はより重要性が増すと予測されるため、更なるスケーラビリティの拡大強化および検索・分析機能の充実化を図る予定である。

また、LogAuditor Mail Saverのアプライアンス・モデルとして、LogAuditor Mail Saver AMを製品化している。アプライアンス・モデルでは、ユーザ規模に応じて、ハードウェアも含め、小規模向け、中規模向け、大規模向けに3タイプにサイジングし簡易な導入を可能としている。

表1. アプライアンス・モデル ラインナップ

モデル	小規模向け	中規模向け	大規模向け
アーカイブ領域	300GB (最大:2.7TB)	1.8TB (最大:7.6TB)	3.6TB (最大:7.6TB)
メール保存期間	1年~3年を想定	1年~3年を想定	1年~2年を想定
ユーザ数	2000ユーザまで	5000ユーザまで	10000ユーザまで

## 参考文献

- (1) Granz, J, F. , 他 : The Expanding Digital Universe, IDC White Paper, (2007)
- (2) 郡光則, 他 : 多種多様なログの統合管理を実現する LogAuditor Enterprise, 三菱電機技報, (2006)
- (3) 郡光則, 他 : 検索機能を備えたストレージシステムによる大規模並列全文検索, 電子情報通信学会技術研究報告, CPSY-2002-47, (2002)