大規模ログデータ検索 I/F の実装

秩父　かおり　　佐藤　重雄　　菅野　幹人
三菱電機株式会社　情報技術総合研究所

1. はじめに

近年、企業は個人情報や企業機密の漏洩防止、内部統制強化を求められ、あらゆる業務の履歴情報をログとして蓄積し分析することによりセキュリティを高めることが重要になっている。

そこで当社では、企業内の各種情報システムで出力される大容量のデータを格納でき、且つ、形式の異なる多種多様なログへの対応が可能なログ専用データベース（ログ DB）を開発している。

本データベースは大容量で多種多様なデータを効率よく蓄積することに主眼をおいたものであり、分析用のアプリケーションは、データベースサーバ上の検索インタフェース（I/F）を用いてシステム毎に作成されていた。しかし、Web 環境での DWH 検索が一般的となった今日においては、ログ DB に対しても Web 環境の汎用的なシステムで検索することが求められている。そのため、ネットワーク経由でクライアントから検索可能にする検索 I/F を新たに開発した。（図 1. システム構成図参照）

本稿では、ログ DB の運用管理機能強化の一環として、ネットワーク経由でクライアントからログ DB の検索を可能とする検索 I/F を実現したので、その実装方式について述べる。

2. クライアントからの検索での要求仕様

ログ DB をクライアントから検索する場合に求められるシステム要件は以下の通りである。

(a) Web 環境からの検索を可能にする。（多数のクライアントに対応可能とする。）
(b) Web サーバでは検索結果を保持せず、画面を表示する都度、検索結果を取得する。
(c) ログ DB の検索結果は 1 レコードが大きなサイズになる可能性があるが、画面に表示しない部分は不要である。
(d) 検索結果を全件取得する前に、1 件分のレコードの全体を取得し、その後最初の検索の結果取得を再開することを可能にする。

Implementation of interface that query large-scale log data
Kaori Chichibu, Shigeo Sato, Mikihito Kanno
Mitsubishi Electric Corporation

検索結果を画面表示する場合、一般的な DWH では、アプリケーションで検索結果を全件取得し、データ表示を制御することが考えられるが、1 レコードのサイズが非常に大きいログ DB では、クライアント上に全ての結果を保持することは通信部分の負荷や、クライアントのリソースの消費を考慮すると効率的ではない。そのため、必要なデータを必要な時に取得できるようにする。

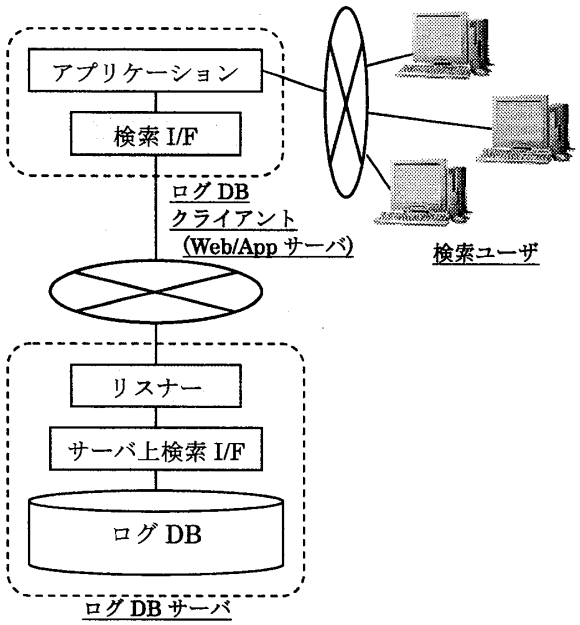

図 1. システム構成図

3. ログ DB の検索 I/F 仕様

クライアントからの検索に利用するデータベースサーバ上の検索 I/F の仕様と、ネットワーク経由で利用する場合の課題は以下である。

(a) 可変長でサイズが非常に大きいデータ型が存在することが特徴のログ DB は、検索実行、検索結果取得、検索クローズの流れで行われる検索時に、大量にデータベースサーバのリソースを消費することがある。
サーバでの検索は限られた人数による検索となるため問題ないが、多数のクライアン

トからアクセスされた場合はリソース消費の抑制が課題となる。

(b) 検索結果を返すバッファは、検索 I/F を利用する側ではなく、データベースサーバが用意する。
可変長でサイズが非常に大きいデータが検索結果に含まれるため、１回の検索で効率的に結果を提示する方式が採られている。
１台のマシン上では問題ないが、ネットワーク経由の場合クライアント側でバッファを確保する必要があり、予め結果サイズがわからないと１レコードの最大サイズを取得件数分確保することになり、不要に大きなサイズのバッファを確保することになる。

(c) ログ DB に対する検索問い合わせ文は、ログの属性項目を用いた検索条件と、ログ本文に対する検索条件で成り立っており、ログ本文の検索条件はファイルを利用して指定する。
ログ本文のサイズが大きい場合検索条件も長くなることが予想されるため、ファイルを利用した指定方法が採られているが、ネットワーク経由の検索の場合は、ファイルではなく通信データとして転送するため、検索 I/F としてもファイルを利用しない指定方法を提供する必要がある。

4. クライアントからの検索 I/F の実装方式

クライアントからの検索を実現するにあたっては、2. で述べた要求仕様を満たし、3. で述べたデータベースサーバでの検索 I/F の課題を解決する必要がある。そのため、以下の実装方式を採用した。

(1) 先読み処理

データベースサーバ側でログ DB を検索のためにオープンしている時間を短くするため、クライアントからの検索結果取得要求とは非同期に、検索結果の先読みを可能とした。先読み後は引き続きクローズ処理を実行することで、データベースサーバのリソースをできるだけ早く解放するようにした。

また、先読みデータは、データベースサーバから返ってきた内容そのままの形式で保持することはせず、通信データ形式でバッファに格納し、管理テーブルで制御する。このように処理することにより、クライアントからの検索結果取得要求時は、バッファから要求件数分の先読みデータを即座にクライアントへ転送可能となり、通信性能も向上する。

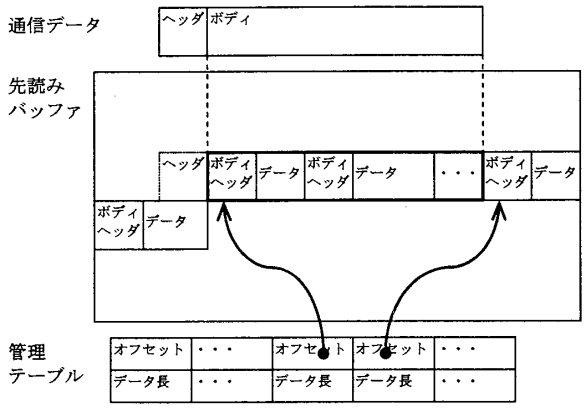

図 2. バッファ管理図

(2) 検索結果のサイズ指定

検索 I/F として、１回に取得する件数とログ本文のサイズを指定可能とし、アプリケーションで適切なサイズのバッファを確保できるようにした。

(3) １レコード明細取得処理

(1)、(2)の処理では、大規模ログデータに対応するため、ログ本文のサイズをアプリケーションで一覧表示可能なサイズに限定した上で、１回に複数件の検索結果を取得することを想定している。一覧表示中に、ログ DB に格納されているデータそのものを取得したい場合に備えて、クライアントからのログ DB の検索実行または検索結果取得中でも、別の１件を取得可能とする機能を提供した。

(4) ログ本文の検索条件の指定

ログ DB の検索問い合わせ文に含まれるログ本文の検索条件は、ログ DB の検索問い合わせ文とは別に、複数の条件を検索 I/F で指定できるようにした。サーバ側でファイルを作成し、ログ DB の検索を実行する。

5. おわりに

本稿では、クライアントからネットワークを介して大規模ログデータを検索するインタフェースの実装方式について述べた。今後は本方式の評価を行い、リソース面と転送性能両面を確認していく予定である。

参考文献

[1]"多種多様なログを統合し、一元管理する"LogAuditor"", 三菱電機技報 2007 年 1 月号「技術の進歩特集」

[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 年 10 月号「情報セキュリティシステム／サービス基盤」