特集論文

クラウド上でのデータベースセキュリティを実現するデータ暗号化機能

佐藤重雄*
山岸義徳**

Data Encryption for Database Security in Cloud

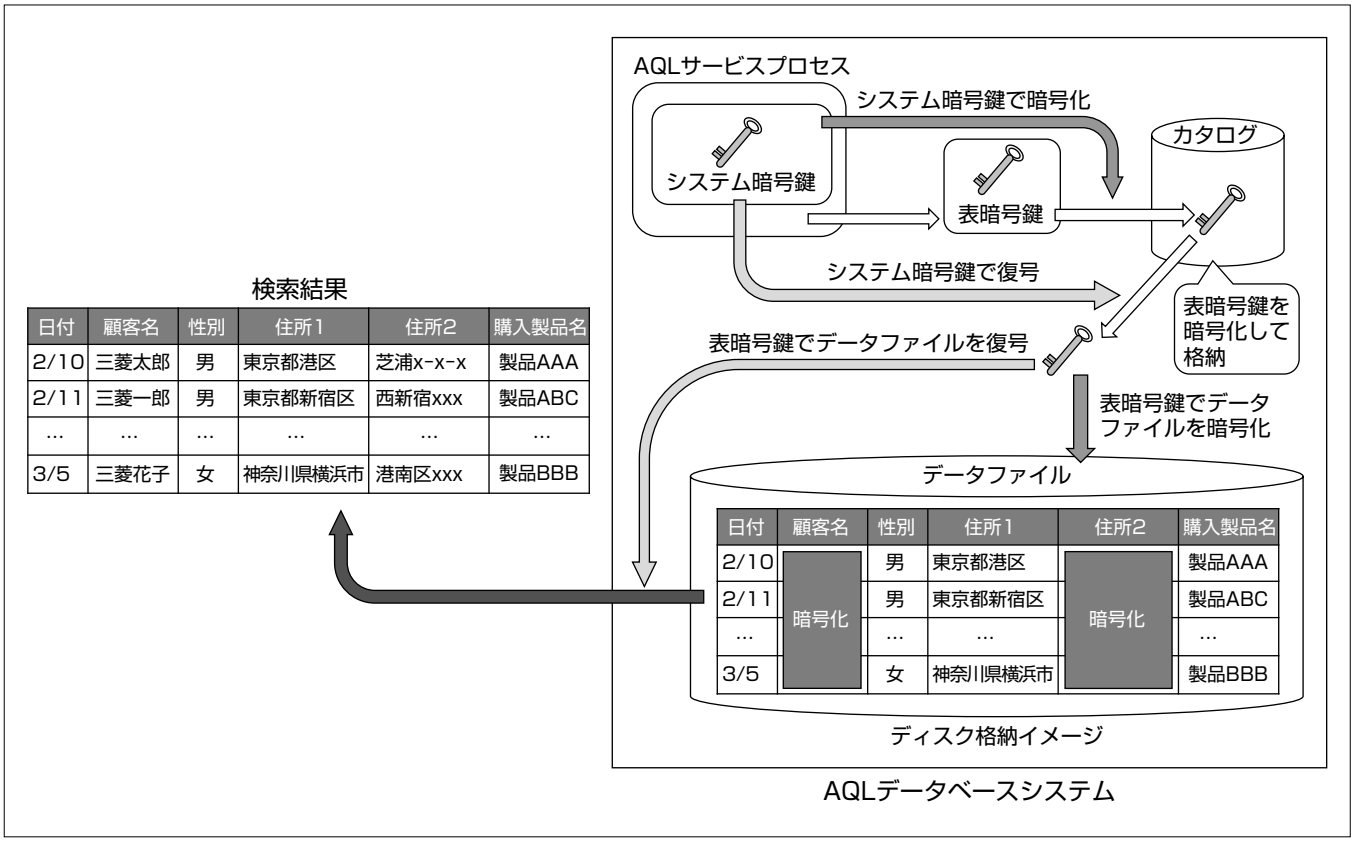
Shigeo Satou, Yoshinori Yamagishi

要　旨

企業が自社製品の販売促進を目的として，顧客から収集した顧客属性を含むデータの分析を行う場合，氏名，住所などの個人情報が分析システムのデータベース内に格納される。近年増加している不正アクセスによるデータ流出の脅威への対応，パブリッククラウドの普及によるクラウド上でのデータベースセキュリティ確保の要求などによって，データベースに対するデータ暗号化の必要性が高まってきた。

三菱電機インフォメーションネットワーク㈱(MIND)が提供するデータ分析フレームワーク“AnalyticMart”は，データ分析を低コストで高速に実現できる製品であり，販売分析，顧客分析などに活用されてきた。AnalyticMartで使用する構造化データベースAQL(Analytical Query Language)では，データベース利用時のユーザー認証機能，表を参照する際のアクセス権限機能を提供してきたが，クラウド環境にデータベースを構築する場合には，システム管理者によるデータベースファイルへの直接アクセスによる情報漏洩(ろうえい)を防止することが課題であった。

AQLで提供するデータ暗号化機能では，個人情報などが格納された特定の項目のみの暗号化が可能であり，二階層で実現する暗号鍵はデータベースシステム内で管理を行うため，暗号化に伴うアプリケーションの変更は不要となる。必要最小限の項目に対する暗号化／復号処理をディスク入出力の前後で並列実行する仕組みを実現したことによって，従来の高速な検索・集計性能を維持したままで，データベースセキュリティを強化することが可能となった。



AQLのデータ暗号化機能

AQLのデータ暗号化機能では，個人情報など特定の項目のみを暗号化して，ディスクに格納することを可能とする。暗号鍵は，データファイルの暗号化／復号を行う表暗号鍵，表暗号鍵の暗号化／復号を行うシステム暗号鍵の二階層で構成しており，どちらの暗号鍵もデータベースシステム内で管理している。

*三菱電機インフォメーションネットワーク㈱　**三菱電機㈱ 情報技術総合研究所

1. まえがき

企業の販売促進活動では，自社商品の販売データと顧客属性とを結びつけ，顧客ごとの購入傾向の分析などが行われるが，顧客から収集した氏名，住所，クレジットカード番号などの個人情報は，分析システムのデータベース内に格納されている。近年，システムの脆弱(ぜいじゃく)性への攻撃による不正アクセスなどによって，収集した個人情報が流失し，企業の信用問題に発展する事故も多くなっている。また，個人情報保護法への対応，パブリッククラウドの普及などによって，データのセキュリティ対策の重要性が高まっている。MINDが提供するデータ分析フレームワークAnalyticMartは，大規模データに対する分析処理の高速化を実現するフレームワークであるが，格納されたデータのセキュリティを強化するためには，データベースシステムでのユーザー認証，アクセス制御に加えて，データファイルへの直接アクセスの脅威に対応することが課題であった。

本稿では，高速な検索処理性能と，データベースセキュリティ強化の両立を実現したAnalyticMartのデータ暗号化機能の特長，実現方式，効果を中心に述べる。

2. AnalyticMartでのセキュリティ対策と課題

2.1 AnalyticMart

AnalyticMartは，販売分析，顧客分析，ログ分析，環境データ分析といった多様で形式の異なるデータの分析を統一したアーキテクチャで効率よく低コストで実現できるフレームワーク[(1)]である。AnalyticMartが提供する統一した分析基盤によって，中小規模から大規模まで，規模に合わせたデータ分析システムの構築・運用を実現することができる。AnalyticMartではAQLと呼ばれる構造化データ用データベースを使用しているが，そのアーキテクチャはデータ分析向けに最適化したものである。AQLの特長[(2)]は次のとおりである。

(1) 高度な圧縮技術によって，元データを1/5から最大1/40に圧縮し，ストレージ容量を大幅に削減することができる[(3)]。

(2) 表データを列単位の形式で格納し，必要なデータの読み出し処理，圧縮データの伸張処理，レコードの選択処理，及び集計処理を複数のプロセッサで並列実行することによって，高速な集計・検索性能を実現する。

(3) データロード用のステージング用サーバと，複数の検索用追加サーバで構成する“検索サーバ追加機能”によって，データ分析システムの利用者が増加した場合に，スケールアウトによる負荷分散を実現できる[(4)]。

2.2 AnalyticMartでのデータベースセキュリティ対策

AnalyticMartのAQLでは，データに対するアクセスを制限する手段として，次の機能を実現している。

(1) データベースシステム接続時のユーザー認証機能

(2) 表の参照権限をユーザー単位，ロール(複数のユーザーを役割でまとめたもの)単位で付与，剥奪する機能

これらの機能によって，データベースシステムを介して表データにアクセスする場合のセキュリティは確保しているが，次のような脅威への対応は十分ではなかった。

(1) システム管理者のデータファイルへの直接アクセス

(2) システムの脆弱性を悪用した外部からのデータファイルへの不正アクセス

(3) ディスクの盗難

また，近年は，パブリッククラウドの普及によって，個人情報や機密データを自社のシステムではなく，クラウド上に格納する場合も多くなり，データファイル自体のセキュリティを確保する手段として，データ暗号化の必要性が高まってきた。

2.3 データ暗号化の課題

データセキュリティを強化するために，データ暗号化を行う際には，次の点が課題となっていた。

(1) 処理性能の低下

データをロードする際の暗号化処理，及び暗号化したデータを検索する際の復号処理のオーバーヘッドによって，処理性能が低下する。

(2) データ容量の増大

あらかじめ暗号化したデータをロードする場合は，データ圧縮の効率が悪くなるため，データベースのファイルサイズが増大する可能性がある。

(3) 暗号鍵の管理

アプリケーション側で暗号化／復号を行う場合は，個々に暗号鍵を管理する必要があり，アプリケーションの変更が必要となる。

AnalyticMartのAQLでは，上記の課題を解決するため，必要な項目(列)のみの暗号化を行うデータ暗号化機能を開発した。この機能では，暗号鍵の管理，データの暗号化／復号は，データベース側で実行し，データロード，検索処理の性能は従来と同等性能を実現した。また，データベースのファイルサイズは，暗号化あり／なしの場合で同一のサイズを実現した。

3. AnayticMartのデータ暗号化機能

3.1 データ暗号化機能

(1) 暗号鍵の生成と管理

AQLのデータ暗号化機能は，暗号化と復号で同じ鍵を使用する共通鍵暗号方式を採用し，次の二階層の暗号鍵を使用する。

①表データの暗号化／復号を行う表暗号鍵

②表暗号鍵の暗号化／復号を行うシステム暗号鍵

これらの暗号鍵はどちらもAQLのデータベース内で管理するため，アプリケーションで暗号鍵の管理を行う必要がなく，暗号化機能を使用する場合もアプリケーションの変更が不要となる。

AQLでの暗号鍵管理を**図1**に示す。表の定義時に，特定の列に対して暗号化属性(ENCRYPT)を設定すると(**図2**)，AQLデータベースシステムは，表単位に異なる表暗号鍵を生成する。表暗号鍵は，データベースの定義情報を格納するカタログ内に格納するが，ディスクの盗難等によって，データベース全体が流出した場合の危険性を考慮し，表暗号鍵はシステム暗号鍵で暗号化して格納する。

システム暗号鍵は，システムごとに異なる暗号鍵であるが，それ自体が静的に保存されることはない。システム導入時にシステム暗号鍵設定コマンドを実行することによって，データベースを実行するサーバ内に，システムごとに異なるシステム固有情報を設定し，AQLデータベースシステムは，起動時にこの情報を基にシステム暗号鍵を生成し，メモリ内に保持する。

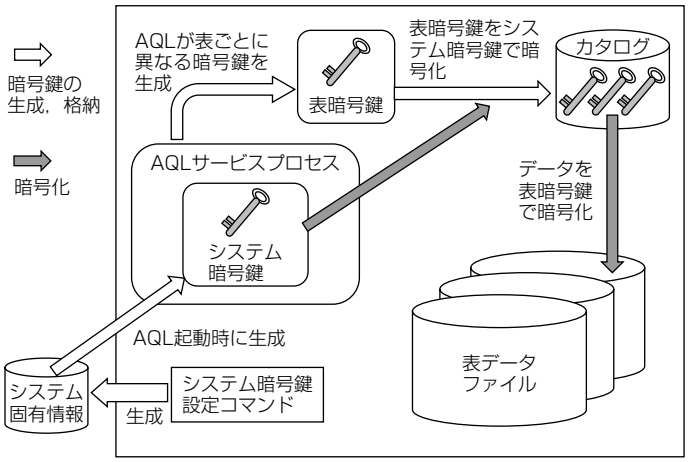


図1．暗号鍵の管理

ディスク格納イメージ

日付	顧客名	性別	住所1	住所2	購入製品名
2/10	暗号化	男	東京都港区	暗号化	製品AAA
2/11		男	東京都新宿区		製品ABC
…		…	…		…
3/5		女	神奈川県横浜市		製品BBB

表定義文

```
CREATE TABLE “購買データ”(
“日付”              DATE,
“顧客名”            CHAR(8)   ENCRYPT,
“性別”              CHAR(2),
“住所1”             CHAR(32),
“住所2”             CHAR(16)  ENCRYPT,
“購入製品名”        CHAR(10),
                ……
);
```

図2．暗号化属性の設定方法

(2) データロード時の暗号化処理

AQLは列単位にデータを格納する構造となっており，列ごとに複数行のデータをまとめた単位(ブロック)で，データの圧縮処理を行った後に，ディスクへの書き込み処理を実行する。暗号化属性が設定された列の場合は，圧縮処理を行ったブロックに対して暗号化処理を実行した後に，ディスクへの書き込み処理を行う(**図3**)。暗号化時には，データベースシステムがカタログから対象とする表の表暗号鍵を取得し，システム暗号鍵で復号した後に，表データを暗号化するための暗号鍵として使用する。なお，暗号化ライブラリはOpenSSLを利用し，暗号化アルゴリズムは，米国標準暗号であるAES(Advanced Encryption Standard)を使用する。

(3) 検索時の復号処理

図3に示すように，検索処理では，必要な列のデータのみをディスクからブロック単位で読み出し，復号処理(暗号化属性が設定されている列の場合)，伸張処理を行った後に，選択，集計などのデータベース処理を実行する。検索時も，データロード時と同様に，カタログから取得した表暗号鍵をシステム暗号鍵で復号した後に，表データを復号するための暗号鍵として使用する。

3.2 データ暗号化機能の特長

AQLのデータ暗号化機能の特長は次のとおりである。

(1) 従来の高速処理性能を維持したままで暗号化を実現

暗号化／復号処理のオーバーヘッドを極力削減し，従来の高速処理性能を維持するために，次の方式を実現した。

①データを列単位に格納するAQLのデータ構造を活用し，個人情報など暗号化が必要な列のみの暗号化を可能とする

②複数の列に対する暗号化／復号処理の並列実行による高速化

③圧縮と暗号化，及び復号と伸張のそれぞれの処理を連続して実行することによって，キャッシュアクセスの効率を向上

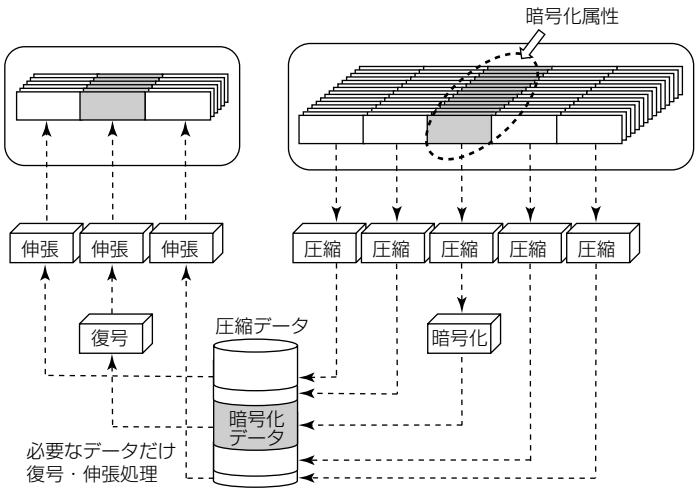


図3．データ暗号化の内部処理

④Intel Xeon(注1)プロセッサの5,600番台以降に搭載されたAES-NI(AES New Instructions)を利用し，AESの暗号化／復号をプロセッサの命令で実行することによる高速化

⑵ 従来と同じデータベースのファイルサイズを実現

AQLでは，列ごとにブロック単位でデータを圧縮してディスクに書き込むが，データ暗号化処理はデータ圧縮後のブロックに対して実行するため，暗号化によってデータの圧縮効率が変わることはない。AQLのデータ格納単位であるブロックのサイズは，暗号化アルゴリズムとして採用したAESのブロック長128Bit(16バイト)の倍数であるため，暗号化でのパディング処理は発生せず，暗号化後のデータベースのファイルサイズは，暗号化なしの場合と全く同じサイズとなる。

⑶ 旧バージョンのデータに対する暗号化を容易に実現

既にAQLを使用してデータベースシステムを構築している場合，この機能を持つバージョンにバージョンアップ後，次の手順によって，旧バージョンでロード済みのデータを暗号化することが可能となる(**図４**)。

①AQLのアーカイブツールを使用して，暗号化対象とする表の全データのアーカイブを生成する。

②対象表のデータを全て削除した後に，特定の列に対して暗号化属性を設定する。

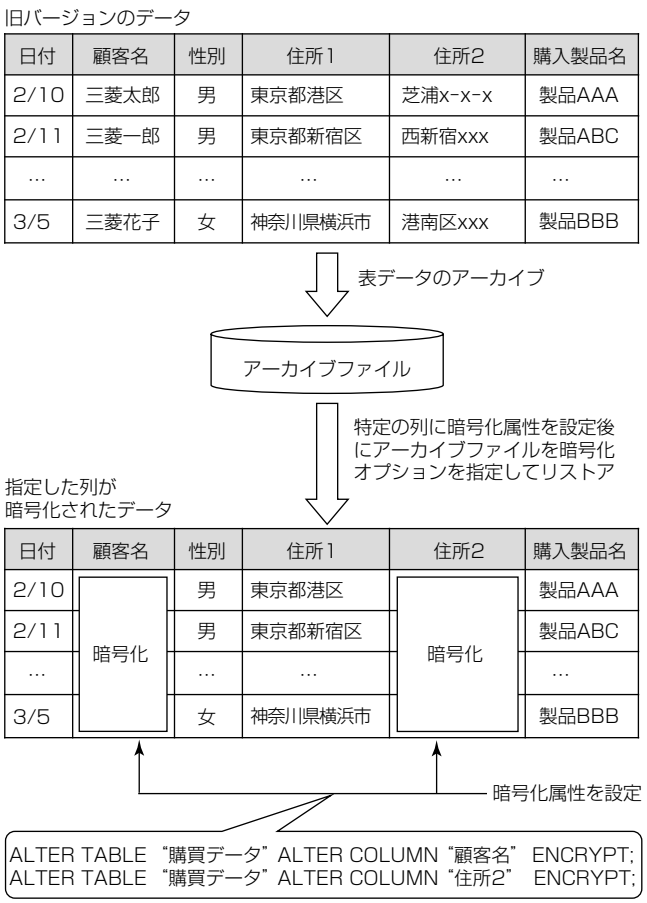


図４．旧バージョンのデータの暗号化

③AQLのアーカイブツールを使用して，①で生成したアーカイブファイルを，暗号化オプションを指定してリストアする。

AQLのアーカイブツールは，圧縮したデータファイルの形式でアーカイブファイルを作成することができる。圧縮処理によって，データファイルは元データの1/5～1/40のサイズになっているため，全データの結果取得，及びデータの再投入を実行するよりも短時間でアーカイブ／リストアが可能となる。アーカイブツールでリストア時に暗号化オプションを設定すると，暗号化属性が設定された列に対しては，アーカイブファイル内のデータを暗号化してリストアするため，データの再ロードを行わずに短時間で特定列の暗号化が可能となる。

(注１) Xeonは，Intel Corp. の登録商標である。

3.3 データ暗号化機能の性能評価

AQLのデータ暗号化機能が，データロード，及び検索処理の性能に及ぼす影響を検証するための評価を実施した。

⑴ 評価モデル

次の表に対するデータロード，検索処理の実行時間を測定した。

①表定義： 25列，レコード長708バイト

②暗号化対象： 15列(レコード長の約50%)

③登録レコード件数： 10億件

⑵ 評価環境

AES-NIの効果を検証するため，AES-NIを搭載したプロセッサと搭載していないプロセッサで評価を実施した。

⑶ 評価結果

データベースのファイルサイズは，暗号化あり／なしの場合で全く同一であることが確認できた。次に処理性能の結果について述べる。

①データロード時間

暗号化なしの表に対するデータロード時間と，AES-NIあり／なしのそれぞれの環境で暗号化を実施した表に対するデータロード時間との相対比を**表１**に示す。AES-NIありの環境では，データ暗号化を行った場合も，暗号化を行わない場合と同等の実行時間であること，AES-NIなしの環境では，データ暗号化による実行時間の増加の割合は５％程度であることが確認できた。

②検索処理時間

暗号化なしの表に対する検索処理時間と，AES-NIあり／なしのそれぞれの環境で暗号化を実施した

表１．データロード時間の相対比

		実行時間(相対比)
暗号化なし		1.00
暗号化あり	AES-NIなし	1.05
	AES-NIあり	1.00

表２．検索処理時間の相対比

			実行時間(相対比)
問合せ１	暗号化なし		1.00
	暗号化あり	AES-NIなし	1.05
		AES-NIあり	1.06
問合せ２	暗号化なし		1.00
	暗号化あり	AES-NIなし	1.09
		AES-NIあり	1.02
問合せ３	暗号化なし		1.00
	暗号化あり	AES-NIなし	1.09
		AES-NIあり	1.01

問合せ1：暗号化対象列15列のうち，６列を取得する検索
問合せ2：暗号化対象列15列のうち，９列を取得する検索
問合せ3：暗号化対象列15列を全て取得する検索

表に対する検索処理時間との相対比を**表２**に示す。暗号化データの復号処理による検索処理時間の増加は，AES-NIありの環境で６％以下，AES-NIなしの環境で９％以下であることが確認できた。また，検索で抽出する列で，暗号化された列が多く含まれるほどAES-NIによる効果が大きくなることも確認できた。

4. むすび

AnalyticMartでのデータ暗号化機能の仕組みと特長などについて述べた。この機能によって，高速な検索処理性能，高圧縮のファイルサイズを維持したままで，データベースのセキュリティを強化することが可能となった。IoT(Internet of Things)の利用拡大によって，GPS(Global Positioning System)情報による個人の行動履歴や，各種センサによる身体情報などより多くの個人情報を収集して分析する機会が増えることが予想される。この機能によって，パブリッククラウド上で個人情報を安全に保管して分析に活用することが可能になるとともに，金融業などデータベースセキュリティの要件が厳しいシステムへの適用も期待できる。今後は，暗号鍵の世代管理，データ移行時の運用の更なる容易化などの機能強化を行っていく。

参考文献

(1) 小出健太，ほか：情報セキュリティを支えるデータ分析フレームワーク“AnalyticMart”，三菱電機技報，**87**，No.7，405～408 (2013)
(2) 山岸義徳，ほか：高速集計検索エンジンとセンサデータベースへの応用，三菱電機技報，**83**，No.12，709～712 (2009)
(3) 郡　光則：データウェアハウス向け高性能データ圧縮方式，情報処理学会論文誌，**47**，No.SIG13，58～73 (2006)
(4) 戎　直哉，ほか：スケールアウトによる負荷分散を実現するデータ分析フレームワーク“AnalyticMart”，三菱電機技報，**89**，No.8，448～451 (2015)