

オープンソース DBMS 拡張によるセンサデータベースの実現

竹内 丈志[†] 山岸 義徳[†] 中村 隆顕[†] 立床 雅司[†] 郡 光則[†]三菱電機株式会社 情報技術総合研究所[†]

1. はじめに

近年、センサや機器が出力するデータを一元管理し、環境・省エネ対策、セキュリティ対策、各種機器の運転管理や障害解析などへの活用ニーズが高まっている。われわれは、膨大なセンサデータを一元管理し、場所別・組織別など多視点でのデータ分析を可能とする大規模・高速なセンサデータベースを開発した。ここでは、オープンソース DBMS (Database Management System) の PostgreSQL を活用することにより、高速性に加えて高機能性も実現し、その有効性を確認した。

2. センサデータベースの方式検討

2.1. センサデータの特徴

センサデータは各種機器が出力する計測点ごとの計測データであり、以下のような特徴をもつ。

- 時系列的に発生する更新を伴わないデータ (追記型データ) である。
- 不定期、あるいは一定間隔で発生し、多数の計測点からミリ秒など短周期で大量に発生する場合があります、データ量が膨大となる。

2.2. RDBMS適用における課題

PostgreSQL のような RDBMS (Relational Database Management System) で大量のセンサデータを管理する場合、次の課題がある。

- (1) トランザクション処理のような特定レコードへの参照・更新処理に適しているが、センサデータのような追記型データに対する集計・検索処理には不向きで処理時間がかかる。
- (2) ストレージ容量が増加し、運用コストが増大する。

2.3. センサデータベースの実現アプローチ

上記課題を解消し、高性能・高機能なセンサデータベースを実現するため、PostgreSQL を拡張し、当社独自の高速集計検索エンジン [1] にセンサデータを格納する方針とした。高速集計検索エンジンは、追記型データの集計・検索に適した

Sensor Database Implementation with Open Source Database Management System and its Expansion

[†] Takeshi Takeuchi, Yoshinori Yamagishi, Takaaki Nakamura, Masashi Tatedoko, Mitsunori Kori
Information Technology R&D Center, Mitsubishi Electric Corporation

データベース[2]である。列単位のデータ取り出しが可能なデータ格納方式とセンサデータ向け圧縮技術[3]により、高速集計を可能とするとともにストレージ容量を概ね 1/5~1/40 に抑えることが可能である。

3. センサデータベースの実現方式

3.1. システム構成

PostgreSQLと高速集計検索エンジンとの連携は、連携機能により実現する (図 1)。また、センサデータ分析アプリケーション (APP) はセンサデータベース API を介してデータの送受信を行う。

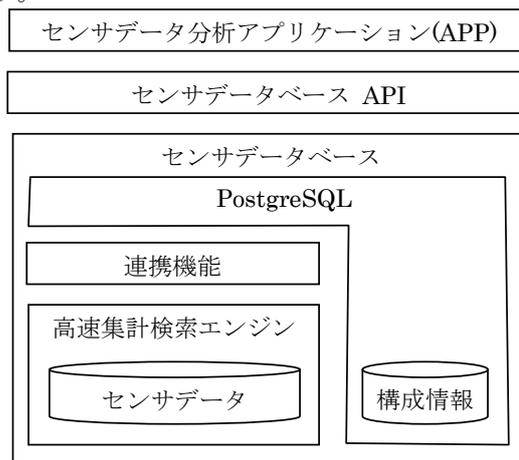


図 1 システム構成

3.2. PostgreSQLの拡張

PostgreSQL の拡張機能を利用することで、PostgreSQL インタフェースによる高速集計検索エンジンへの集計処理を実現する。

SQL の拡張構文を指定することにより、連携機能を介して高速集計検索エンジンに集計・検索の問い合わせを発行することが可能である。また、高速集計検索エンジンから得られる結果セットは、連携機能により PostgreSQL に提供可能なデータ形式へ変換される。PostgreSQL はこの結果セットに対して、演算、結合、ソート等の任意のデータ操作を行うことが可能である。

3.3. データベース設計

ここでは店舗ごとの電力消費量の例を図示し

た図 2を例に、センサデータベースの構造について説明する。図 2のように場所別・組織別など多視点でデータ分析には、階層関係や各計測点との関係定義に関する情報（以下、構成情報）の管理が必要である。この構成情報の管理は PostgreSQLで行う。また、計測点ごとのセンサデータは列方向に計測点を配置するようなテーブル構造を定義し、高速集計検索エンジンで管理する。このようなテーブル構造により、計測点ごとの集計に対して列単位の効率的な処理を可能とする。

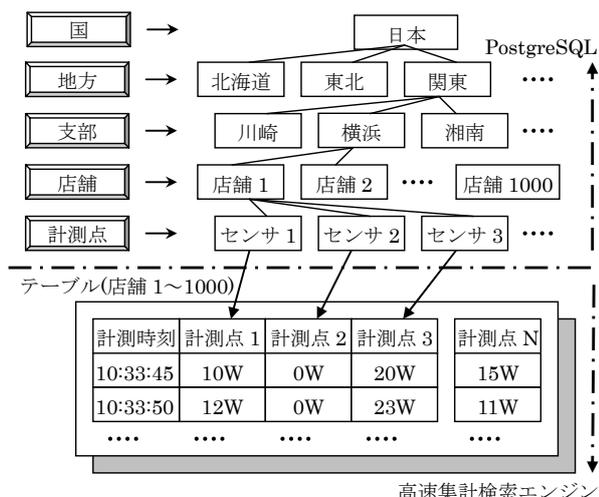


図 2 データベース構造の概念図

3.4. 集計・検索の処理手順

本センサデータベースは、PostgreSQLと高速集計検索エンジンを連携動作させることにより、APPからのデータ分析要求を処理する。図 2に示すデータベース構造の場合、センサデータの集計・検索するためには以下に示す 2 個の検索クエリによる 4 段階の処理で実現する（図 3）。図 2を例に、横浜支部の店舗ごとの電力消費量を集計する場合の処理手順を以下に示す。

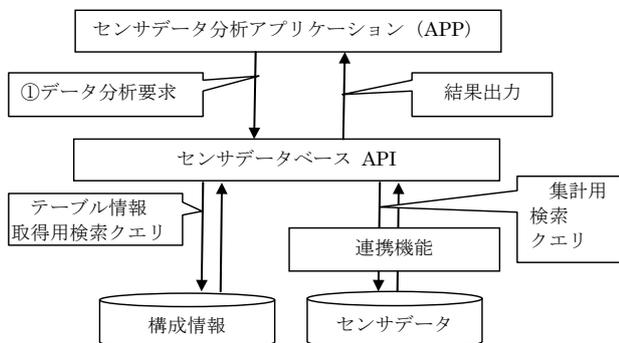


図 3 集計結果取得までの処理の流れ

横浜支部各店舗の電力消費量の集計要求を受付け
 横浜支部の各店舗に関連付く計測点が高速

集計検索エンジンのどのテーブルのどの列に対応づくか、構成情報を検索してテーブル情報を取得

の結果から電力消費量を店舗ごとに集計する検索クエリを生成し、PostgreSQL から店舗ごとの集計結果を取得

横浜支部各店舗の電力消費量の集計結果を出力

以上により、PostgreSQL のインタフェースから高速集計検索エンジンを利用可能となり、2 個の検索クエリ生成と実行による計測点ごとあるいは多視点のセンサデータの集計を可能とした。

4. 評価とまとめ

本センサデータベースを実現し、高速集計検索エンジンとの連携動作、ならびに、表 1の条件で性能評価を実施した（表 2）。PostgreSQL単体よりも約 20 倍高速な集計性能と、94.7%の圧縮率(*)を示し、本データベースの有効性について確認した。

* (圧縮前サイズ-圧縮後サイズ) / 圧縮前サイズ

表 1 評価環境

データ規模	3.25GB 10,765,000 件/計測点 (77 計測点、828,905,000 件)
集計条件	特定拠点の 10 計測点を集計
PC サーバ	CPU : Core 2 Quad 3.00GHz MEMORY : 8GB HDD : 1TB, 160GB (共に 7,200rpm) OS : Windows Server 2008 R2

表 2 評価結果

	集計時間		DB サイズ	
	[sec]	相対比	[GB]	相対比
本センサデータベース	1.701	1.0	0.17	1.0
PostgreSQL 単体	33.711	19.8	4.32	25.4
高速集計検索エンジン単体	1.653	0.9	0.17	1.0

参考文献

- [1] 山岸義徳, 他: 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, Vol.83, No.12, pp.11-14(2009)
- [2] 郡光則, 他: 高性能並列情報検索技術, 三菱電機技報, Vol.83, No.12, pp.7-10(2009)
- [3] 加藤守, 他: 環境情報データベース向け高性能センサデータ圧縮方式, 情報処理学会第 73 回全国大会