

時系列センサデータシステム TIVA の開発

—シンプル・コンパクトで高速な、時系列データのためのデータベース—

1. 背景

センサネットワーク技術の発達によって、実空間の多様な情報を収集することが可能となった。温度や照度などの環境情報だけではなく、携帯電話に組み込まれたセンサからユーザの位置や加速度などの活動情報を取得することも可能である。同時に、取得したセンサデータを利用したこれまでにないサービスやアプリケーションが期待されている。取得可能なセンサ情報を蓄積し他のセンサ情報と組み合わせることで、多様な活動情報を抽出し様々なサービスを実現することが出来る。また、Google や Amazon が Web 上の多くの情報を集めることで新たなサービスや価値を生み出したように、今後は実空間の取るに足らない情報でも大量に集めることで新たなサービスや価値が生まれる時代が来ると期待している。

2. 目的

本プロジェクトでは、過去のセンサデータの解析や参照、そして過去のセンサデータとの比較を可能とするミドルウェアとして、大量のセンサデータの蓄積と高速な検索・参照を可能とするセンサデータベースシステムの開発を目的とする。また、センサデータを用いたアプリケーションの開発を行う一般ユーザに利用されるセンサデータベースシステムとして、センサデータを用いたアプリケーション、サービス、ビジネスの創出を促進することを目指す。

TIVA を利用することで、例えば設置したセンサから取得したデータを過去まで遡ってウェブから参照できるようなシステムを簡単に構築できるようになる。また、さまざまなセンサデータを組み合わせて利用することで、多様な観点からユーザの過去を参照できるライフログのようなサービスや、過去の行動履歴からユーザにあったダイエット方法を提案するヘルスケアサービスなどが実現可能となる。主な利用者は、研究者や学生、スタートアップのサービスに関わるベンチャーなど、コストをかけずにすぐに導入できる環境を必要としているユーザを想定している。

3. 開発の内容

本プロジェクトでは、全てのセンサデータを蓄積して、そのデータをアプリケーションから利用可能とするセンサデータベースシステム TIVA を DBM として実装した。TIVA では、図1のようなシンプルなデータ構造を内部に持ち、カラムストアの技術を用いた効率のよいデータ圧縮を実現するとともに、Sparse Index と呼ばれる索引付けや、新規に到達したデータの非同期更新を行うことでデータの挿入・参照の高速化を実現した。また、TIVA では SQL に慣れた多くのユーザにとって使いやすいデータベースとして、SQL ライクなクエリを実現した。TIVA のプログラムは全て C 言語で実装されており、インストールには gcc 等の C コンパイラや C 言語の標準的なライブラリ群が必要である。Xcode インストール済みの MacOSX10.6 と build-essential インストール済みの Ubuntu Linux8.04 環境で動作の確認をしている。

TIVA では、図1のようなシンプルなデータ構造に時系列なセンサデータを蓄積し、データの時系列性を利用した索引付けや圧縮技術を利用することで既存のデータベースより高速な動作を実現した。

Time	value1	value2	value3
2010/7/30 10:00:00	0.94	0.52	0.73
2010/7/30 10:00:01	0.88	0.42	0.83
2010/7/30 10:00:02	0.82	0.49	0.91
2010/7/30 10:00:03	0.69	0.39	0.81
2010/7/30 10:00:04	0.19	0.56	0.87
2010/7/30 10:00:05	0.17	0.59	0.78
2010/7/30 10:00:06	0.10	0.44	0.74
2010/7/30 10:00:07	0.14	0.47	0.73

図 1. TIVA のデータ構造

また、TIVA はシンプルな DBM のインターフェースとともに、SQL ライクなクエリを実現する。TIVA を利用したプログラムのサンプルを図2に示す。ユーザは、使い慣れた SQL ライクなクエリによりプログラムを記述することができる。

```

#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <tiva/tiva_storage.h>
#include <tiva/tiva_sql.h>

int main()
{
    TIVADB *tdb = tiva_new();
    char time_str[128];

    tiva_init(tdb, "sampledb", INT, 3);

    if( !tiva_query(tdb, "insert into sampledb values('2010/07/31 10:00:00', 10, 100, 1000)") ){
        printf("%s\n", tdb->emsg);
    }

    // get 1 record
    if( !tiva_query(tdb, "select from sampledb where time = '2010/07/31 10:00:00'") ){
        printf("%s\n", tdb->emsg);
    }else{
        printf("get 1 record\n");
        TIVA_ROW row = tiva_store_record(tdb);

        timeval2str((struct timeval*)row[0], time_str);
        printf("%s:", time_str);

        int i;
        for(i=1; i!= 4; i++){
            printf(" %d", *(int*)row[i]);
        }
    }
    printf("\n");

    tiva_close(tdb);
    return 0;
}

```

図 2. TIVA を利用したプログラム例

4. 従来の技術(または機能)との相違

TIVA ではシンプルなデータ構造を内部に持ち、カラムストアの技術を用いた効率のよいデータ圧縮を実現するとともに、Sparse Index と呼ばれる索引付けや、新規に到達したデータの非同期更新を行うことでデータの挿入・参照の高速化を実現した。実際に既存のデータベースとの性能比較を行ったところ、既存のデータベースシステムである Tokyo Cabinet や MySQL より高速に動作することが確認された。

また、TIVA では SQL に慣れた多くのユーザにとって使いやすいデータベースとして、SQL ライクなクエリを実現した。DBM の多くはシンプルな関数群しかサポートしないが、SQL をサポートすることで SQL に慣れたユーザにとって利用しやすいデータベースシステムを目指した。

5. 期待される効果

本プロジェクトで開発した TIVA を利用することで、センサデータを扱うユーザにとって、高速で使いやすいデータベースを提供することが可能となる。まずは、センサデータや実空間情報を利用する技術者・研究者・開発者を対象としてユーザや開発者を獲得し、そして今後の実空間情報を用いたサービスやアプリケーションの創出に貢献したい。

6. 普及(または活用)の見通し

今後、データの圧縮をさらに工夫するとともにセンサデータの利用シーンに合わせた機能(イベント情報による索引付けと検索・分散化など)を充実させることで既存のシステムと更なる差別化を進めることを考えている。同時に、実空間情報を利用したサービスに関わって、その基盤として本システムを利用してもらうことを目指す。

サービス基盤の技術として実績を出して、またソースコードをオープンソースとして公開することで、更なる利用者・開発者を増やすことを目指したいと考えている。

7. クリエータ名(所属)

高木潤一郎(東京大学先端科学技術研究センター森川研究室)

(参考)関連URL

<http://tjun.jp>
