Design and Implementation of Data Management Scheme to Enable Efficient Analysis of Sensing Data

Yuichi Hashi†, Kazuyoshi Matsumoto†, Yoshinori Seki†, Masahiro Hiji‡, Toru Abe‡, Takuo Suganuma‡
†Hitachi Solutions East Japan, Ltd.
‡Tohoku University

Self-IoT 2015, 7 July 2015, Grenoble, France
Contents

1) Introduction
2) Community Environment Sensing Platform
3) Data Management Scheme
4) Use Case scenario
5) Conclusion
Introduction

- Smart Community
  - Build efficient and sustainable social infrastructure

- Sensing data (big data) analysis
  - Current status
    - Focus on analysis of sensing data
    - Depend on Volume of the data and Variety is insufficient
      - Limit exists to the acquisition of useful knowledge
  - Only sensing data is not enough
    - It is necessary to analyze sensing data in conjunction with information of human activities (social context)
    - Social context widens Variety, new knowledge will be obtained
Volume and Variety

Sensing Data: human activities, environment

Event Data: repair, ownership

Social Context: .................

Variety
CESP – Community Environment Sensing Platform for community-care

- CESP is a platform to monitor a variety of statuses of each element in a smart community to manage and analyze the monitored data for the benefit of the community.
Ownership and Consent Agreement

- Data is collected from various owners
- Owners of each data are different
- Each owner sets consents for use of data individually
Data managed in CESP and its features

- Sensing data
  - Data measured by the sensors of IoT devices installed in a community
  - The size of each Sensing data is small, but a large volume of data is accumulated over time

- Data source information (meta-data)
  - Information about the instrument that generates data
  - The size of Data source information is small and is rarely updated

- Event information
  - Information of events that affects the usage and interpretation of the collected data at the time of data analysis
  - The size of Event information is small and is rarely updated
### Design of application

- **Conventional application**
  - Owner of Application = Owner of Data
    - Owner needs to prepare the Data for Application to use

- **Current application (Cloud application)**
  - Owner of Application ≠ Owner of Data
  - Data from various owners
    - each owner sets consents for use of data individually
  - Application should use the data in accordance with the intention of each owner

![Diagram showing the flow of 'consents for use' between the cloud and a user's computer.]
Intelligent Knowledge-as-a-Service
To achieve both high-speed search of large volumes of data, and flexible search of data/information for data analysis and application knowledge acquisition.

It combines a schema-free, Document-type database and a Graph-type database suited for flexible search.
Dataset and Event

Dataset

Key.1-value.1
Kea.2-value.2

Create New Dataset

Key.k-value.k
Key.l-value.l

Sensor

sensing

Modify
Repair
Change policy

Event

sensing
Intelligent Knowledge-as-a-Service

CESP Data Model

- **Storage**
  - :Own (Owner)
  - :ConnectTo (Gateway)

- **Gateway**
  - :Own (Owner)
  - :ConnectTo (Composite Data Source)

- **Composite Data Source**
  - :ComposeOf (Dataset)

- **Dataset**
  - :Measure (Data Source)

- **Data Source**
  - :KindOf (Location)

- **Location**
  - :HasAccountOf (User)

- **User**
  - :Own (Owner Event)
  - :ConnectTo (Environment)

- **Environment**
  - :TargetOf (Thing)

- **Thing**
  - :TargetOf (User)

- **Owner Event**
  - :Own (Owner)

- **Event information**
  - Owner Event
  - Environment

- **Source Event**
  - :HasAccountOf (Source Event)

- **Target Event**
  - :TargetOf (Dataset)

- **Dataset**
  - :KindOf (Source Event)

- **Social Context**

- **Sensing Data**
  - Dataset Fragment
  - Data Content
### Type of Event Information – sensor/device

<table>
<thead>
<tr>
<th>Devices</th>
<th>type</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>device</td>
<td>Installed</td>
<td>new installation of a device</td>
</tr>
<tr>
<td></td>
<td>Modified</td>
<td>modification of device parameters</td>
</tr>
<tr>
<td></td>
<td>Repaired</td>
<td>repair of a device</td>
</tr>
<tr>
<td></td>
<td>Stopped</td>
<td>stop of use of a device</td>
</tr>
<tr>
<td>gateway</td>
<td>Installed</td>
<td>new installation of a gateway</td>
</tr>
<tr>
<td></td>
<td>Modified</td>
<td>modification of gateway parameters</td>
</tr>
<tr>
<td></td>
<td>Repaired</td>
<td>repair of a gateway</td>
</tr>
<tr>
<td></td>
<td>Stopped</td>
<td>stop of use of a gateway</td>
</tr>
<tr>
<td></td>
<td>Change Route</td>
<td>Change of the route of a device to a gateway</td>
</tr>
</tbody>
</table>
### Type of Event Information – owner/things

<table>
<thead>
<tr>
<th>target</th>
<th>type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ownership</td>
<td>Set</td>
<td>set ownership to device</td>
</tr>
<tr>
<td></td>
<td>Changed</td>
<td>change owner</td>
</tr>
<tr>
<td>policy</td>
<td>Set</td>
<td>set policy to device</td>
</tr>
<tr>
<td></td>
<td>Changed</td>
<td>change policy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>target</th>
<th>type</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>things</td>
<td>Installed</td>
<td>new installation of thing</td>
</tr>
<tr>
<td></td>
<td>Repaired</td>
<td>repair of thing</td>
</tr>
<tr>
<td></td>
<td>Broken</td>
<td>broken of thing</td>
</tr>
</tbody>
</table>
- Document-type DB
  - CouchDB
- Graph-type DB
  - Neo4j
- Query Function by Java/JavaScript
  - Dispatch Query to CouchDB or Neo4j
  - Rewrite parameter in query (automatic range calibration)
Relationship between Dataset, Data source information and Event information

Sensor

beAffected

ref:Seq

Output

Sensor

ref:Seq

Output

Dataset.current

Dataset.old

beAffected

ref:Seq

Output

ref:Seq

Output

http://.../Dataset.current

http://.../Dataset.old

Event.k+1

description

contents

Event.k

description

contents

interval

60

http://.../Dataset.current

http://.../Dataset.old

http://.../Dataset.current

http://.../Dataset.old

installed

kind

before

after
Graph representation of Data source information

- **Storage**
  - ConnectTo **Gateway**
- **Gateway**
  - ConnectTo **Device**
- **Device**
  - ComposedOf **ref:Seq**
- **ref:Seq**
  - rdf: 1 **Sensor.1**
  - rdf: 2 **Sensor.2**
- **Sensor.1**
  - kind **Thermo**
- **Sensor.2**
  - kind **CO2**
- **hasSpec.** **Specification**
- **Installed.** **GeoLocation**
- **Own.** **Owner**
- **type** **Opened**
- **type** **Noopened**
- **Community.A**
- **measure**
**Performance Measurement Environment**

Sensors collect data every 5 minutes ⇒ 288 records by one sensor per day

<table>
<thead>
<tr>
<th></th>
<th>Client</th>
<th>Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Core i3-1.70GHz, 2Core/4thread</td>
<td>Core i7-3.9Ghz, 4Core/8Thread</td>
</tr>
<tr>
<td>Memory</td>
<td>8GB</td>
<td>8GB</td>
</tr>
<tr>
<td>Disk</td>
<td>256GB SDD</td>
<td>256GB SDD</td>
</tr>
<tr>
<td>OS</td>
<td>Windows 7 Pro. 64bit</td>
<td>CentOS Linux 7.0.1406 64bit</td>
</tr>
</tbody>
</table>
### Measured Times

#### Search time (msec)

<table>
<thead>
<tr>
<th>Position of search data</th>
<th>record size n (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28,800</td>
</tr>
<tr>
<td>1st</td>
<td>24.0</td>
</tr>
<tr>
<td>n/2-th</td>
<td>24.0</td>
</tr>
<tr>
<td>n-th</td>
<td>23.0</td>
</tr>
</tbody>
</table>

#### Registration time (msec)

<table>
<thead>
<tr>
<th>registered record size (B)</th>
<th>record size (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2,880</td>
</tr>
<tr>
<td>288</td>
<td>233.0</td>
</tr>
<tr>
<td>2,880</td>
<td>262.0</td>
</tr>
<tr>
<td>28,800</td>
<td>665.0</td>
</tr>
</tbody>
</table>
Use case scenario of Data Management Scheme

- Validation of Data Management Scheme
  - Flexibility to process various data query patterns
  - Performance to retrieve the data
  - Effectiveness of Event information to the analysis

- Use case scenario
  - Town Management Service
  - Health Support Service
Forecast amount of snowfall at a specific geolocation

- Mr. S noticed that snowfall is predicted for tomorrow morning. He executes the community support application. The application fetches the real-time environmental sensor data (temperature, amount of snowfall) and that of last year from the database. From analysis of the data, the amount of snowfall of tomorrow morning is predicted and the work plan of snow shoveling is represented.

```
QueryDataSource(Location="T-area" & kind = “Temp” & kind = “snowfall”)
```

```
QueryData(Sensor IDs, StartDate=2014/12/10, EndDate=2015/3/20)
```

Analyze the sensing data and forecast amount of snowfall
Propose a data management scheme to realize CESP
  - high-speed search of large volumes of data
  - flexible search of data/information

Measure performance
  - Search time is independent of the data size and position of the data in CouchDB

Show several use case scenarios in a community
  - Town management service
  - health support service

Future work
  - implement a data management scheme and evaluate performance using various use case scenarios
The work is supported by the collaborative European Union and Ministry of Internal Affairs and Communication, Japan, Research and Innovation action: iKaaS. EU Grant number 643262.
This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 643262.