SeaCloudDM: Massive Heterogeneous Sensor Data Management in the Internet of Things

Jiajie Xu

Institute of Software, Chinese Academy of Sciences (ISCAS)

2012-05-15
1. Challenges in IoT Data Management
2. SeaCloudDM: Architectures and Solutions
3. Conclusion
1. Challenges in IoT Data Management

2. SeaCloudDM: Architecture and Solutions

3. Conclusion
Internet of Things:

The Internet of Things (IoT) refers to uniquely identifiable objects (things) and their virtual representations in an Internet-like structure.
An IoT system may contain various kinds of sensors whose sampling data have *heterogeneous* data structures (heterogeneity).

For instance, traffic sensors:

- GPS sensors
- RFID readers
- video-based traffic-flow analysis sensors,
- traffic loop sensors
- road condition sensors
- ....
The data to be managed in IoT are \textit{massive, dynamic} data

- \textbf{Multimedia Data} from video cameras, telemetric devices, and other similar devices;

- \textbf{Sensor data} are sampled frequently, resulting in large data size;

- \textbf{Sensor data} are dynamically change data streams with arriving of new data and deleting of old data.
**Spatial-temporal attribute** is intrinsic for IoT data (Spatial-Temporal Logic)

- Queries cannot be answered through keyword match with SensorID. Instead, through *Spatial-Temporal Constrains*;

- Queries cannot be answered through keyword match with Time-stamps. Instead, through *Spatial-Temporal interpolations*. 
1. Research Background

2. SeaCloudDM: Architectures and Solutions

3. Conclusion
Architecture of Sea-Cloud-based massive heterogeneous sensor Data Management (SeaCloudDM) framework

Sea-Computing: Sensor Connection & Raw Sampling Data Processing

Sea-Computing Node

Sea-Computing Node

Sea-Computing Node

Sea-Computing Node

Sea-Computing Node

Sensor Deployment Layer

traffic sensors

hydrological sensors

geological sensors

video analysis sensors

telemetric analysis sensors

moving video analysis sensors

Cloud Data Management Layer

RDB-KV Cloud: Relational DB and Key-Value Combined Cloud Storage

RDB-KV DB: Unified DB for Heterogeneous Sensor Data Management

Data Analysis & Application Layer

OLAP, Statistical Analysis & Data Mining based on Massive Sensor Data

Information Retrieval and Intelligent Recommendation

IoT Applications

(ITS, Smart-Grid, Emergency Management…)

Sea-Cloud-based Cooperative Data Management Mechanism
1. Sea-Data Processing Mechanism

- Heterogeneous input from various kinds of sensors, standardized (uniform) output data to Cloud Data management layer;

  \[ \text{SamplingValue} = (t, (x, y), npos, \text{schema}, \text{value}) \]
  \[ \text{SamplingSequence}=(\text{schema}, (t, ((x, y), npos, \text{value}, \text{flag}))) \]

- Multimedia data analysis: multimedia data → numerical data;

- Key samplings identification and extraction

  State-threshold based extraction method.
2. Cloud-Data Management Mechanism

- **Relational DataBase and Key-Value store combined (RDB-KV) model**
  - **Homogeneous** database nodes for **heterogeneous** sensor sampling data (both historical and present);
  - DBMS-kernel based data types, operators and indices for sensor data management, supporting both **SQL** and **keyword** search;
  - Sensor Sampling Sequence: for data stream and for **spatial-temporal interpolation** query processing.
RDB-KV Database for Uniformed Storage of Heterogeneous Sensor Data

Example: Data Types, Operators for sensor sampling data

Data Types

- Spatial
- Temporal

Queries

- Spatial-Temporal
- Keyword

Operators

| SamplingValue = (t, (x, y), npos, schema, value) |
| SamplingComponent = (cSchema, cValue) |
| SamplingSequence = (schema, (t, ((x, y), npos, value, flag))_{j=1}^{n} ) |

Create Table IoTData

| (ObjectID: String, ObjectType: String, Owner: String, DeployedTime: Instant, Samplings: SamplingSequence) |

SELECT ObjectID, ObjectType, Owner, DeployedTime FROM IoTData
WHERE ObjectType="Temperature" AND inside(getLocation(Samplings), region) AND
getComponent(atInstant(Samplings,d),1)>50;

SELECT keySearch("BeijingTraffic").Samplings FROM IoTData;

atInstant:
SamplingSequence × Instant → SamplingValue

samplingAppend:
SamplingSequence × SamplingValue × Bool → SamplingSequence

getInstant: SamplingValue → Instant

getPosition: SamplingValue → Point

getNetPosition: SamplingValue → String

getComponent:
SamplingValue × integer → SamplingComponent

= : a × β → Bool

where a, β ∈ (samplingComponent) ∪ BASE ∪ SPATIAL.

keySearch: string → set(tuple)
Sensor-Sampling-Sequence Spatial-Temporal Tree (S4T-Tree)

- Spatial R-Tree Index ➜ static objects
- Grid-Sketched Spatial-Temporal R-Tree Index ➜ moving objects

RDB + KV Database Engine:

- Keyword-Tuple Reverted-File Index ➜ keySearch Operator
RDB-KV Cloud for Management of Massive Sensor Sampling Data

- **Geo-based Data distribution:**
  - Sampling sequence truncation;
  - Home (registered) location based distribution (Hilbert Curve Symbolization).

- **Global Indices: Spatial-Temporal, Keyword index**
1. Research Background
2. SeaCloudDM: Architectures and Solutions
3. Conclusion
**Conclusion**

- **SeaCloudDM framework** with a sea-cloud-based cooperative data management mechanism is proposed;
  - **Sea (70%~90%) + Cloud (10%~30%) Data Management**

- A RDB-KV database model is proposed, with related data types, operators, and indices defined;
  - manages heterogeneous sensor sampling data in a uniformed manner

- A RDB-KV cloud data management model is proposed to manage massive sensor sampling data.
  - combines the advantages of both relational databases and key-value stores
Thanks!