netidee

STIPENDIEN

Data Management Strategies for Near Real-Time Edge Analytics

Zwischenbericht | Call 13 | Stipendium ID 3793

Inhalt

1 Einleitung

The number of IoT sensors and devices is rapidly increasing, and consequently the amount of constantly generated data. This poses crucial challenge to investigate novel and efficient data management strategies at the edge of the network, close to data producers. However, unlike distant and resource-rich cloud data centers, edge nodes (known as cloudlets, edge servers, micro data centers), located close to the source of data, contain limited resource capabilities. In this thesis, therefore, I aim to identify data management strategies towards architecturing of novel edge data services, while coping with challenges such as incomplete sensor data, limited storage capacities and handling only most relevant data. At the same time, I aim to ensure required accurate and timely decisions for emerging IoT applications (e.g., smart buildings, e-Health, intelligent traffic monitoring). From the design viewpoint, I expect to identify which data should be kept at the edge, how long they should be stored, and which approaches can assist in solving these problems.

2 Status

2.1 Meilenstein 1 (2018Q4) – Identification of Requirements for Edge Analytics

Kurzbeschreibung der Haupttätigkeiten
Erkenntnisse zur Vorgangsweise
Kurzbeschreibung der erreichten Ergebnisse
Besondere Erfolge/ Probleme
Gab es große Abweichungen zum Plan? Warum?

To identify requirements for edge analytics the proper literature review is performed. It resulted in a comprehensive overview of data analytics requirements at the network edge from several scenarios and emerging IoT applications that demand near real-time decisions. After investigating state-of-the-art, necessary methodologies (such as edge computing, statistical analysis and data mining) are determined in order to be further explored in solving the main objectives in this work.

One of the main difficulties that researchers often face in their research studies, is to find a proper source of relevant datasets and traces for later experimental evaluation of the proposed approaches. However, several sources are found and can be used later in simulations. Besides that, there were no significant obstacles or deviations from planned activities for this time period.

2.2 Meilenstein 2 (2019Q1) – Edge Data Management Framework (EDMFrame) and Architecture Definition

Kurzbeschreibung der Haupttätigkeiten
Erkenntnisse zur Vorgangsweise

Kurzbeschreibung der erreichten Ergebnisse
Besondere Erfolge/ Probleme
Gab es große Abweichungen zum Plan? Warum?

Based on previous survey in which state-of-the-art works lack contributions in efficient data management strategies for edge analytics, for this milestone, it was necessary to identify and describe in details necessary elements of edge mechanism (such as modules and components) that should lead to efficient data-driven decisions at the edge. Firstly, all fundamental elements (Figure 1) of EDMFrame are defined. Secondly, we proposed an approach to automatize edge data recovery of incomplete data, that is, a multiple-technique recovery with Projection Recovery Maps (PRMs) that detects an optimal trade-off between the gap size (number of missing values) and a range of historical data necessary to keep at the edge (maintained by the *mediator* component). Finally, we submitted a manuscript about the results and achievements to an IEEE journal paper. Once accepted, a preprint version or link details will appear on the project website https://www.netidee.at/data-management-strategies-near-real-time-edge-analytics).

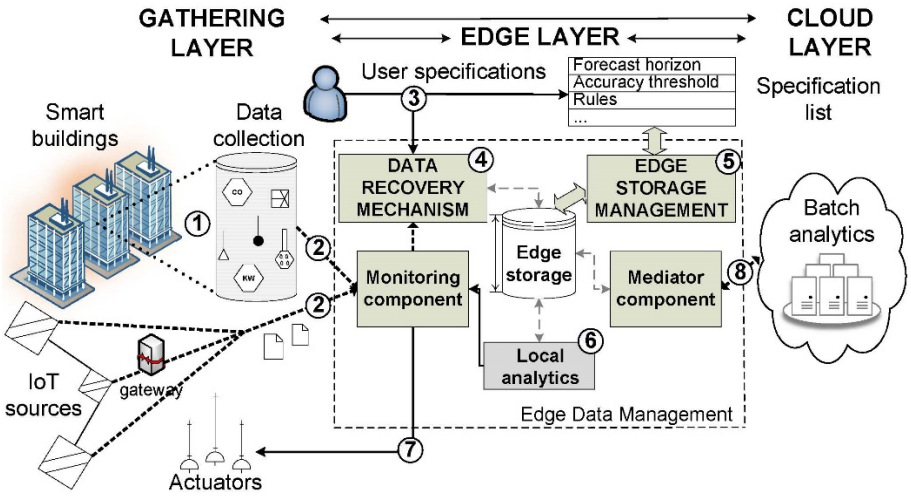

Fig. 1. EDMFrame high-level architecture overview

As explained in Section 3, further investigation on approximate data representation (scheduled for this quarter), potentially leading to more efficient data transfer and edge storage, is postponed to the following time period (2019Q3).

2.3 Meilenstein 3 (2019Q2) – Quality-aware Edge Data Analytics Model

Kurzbeschreibung der Haupttätigkeiten
Erkenntnisse zur Vorgangsweise
Kurzbeschreibung der erreichten Ergebnisse
Besondere Erfolge/ Probleme

Gab es große Abweichungen zum Plan? Warum?

In this time period, elasticity and scalability of edge data storage nodes have been analyzed in order to propose engineering principles for quality-aware edge data services. Optimized design for the storage service management depends on three aspects, namely

- data/system characterization (for example, different metrics and properties);
- edge system/service operations (for example, elasticity and data quality management) and
- application-specific context with needed processing utilities (for example, data approximation and predictive analytics).

From the analysis results, we identified most relevant examples of data/system characteristics (Table 1) and application contexts (Table 2) as well as their impact on the architectural design and engineering of edge elastic storage services. Further, metrics for end-to-end monitoring of elastic storage services are detected through four stages: data collection, data preprocessing, edge storage services and data analytics. Finally, the findings are summarized in a research paper and together with proposed engineering principles are planned to be published in conference proceedings.

Table 1 Data/system characteristics and impact on architectural design

Characteristics	**Impact on architectural design**
Incomplete data	Efficient and adaptive *data recovery* mechanisms;
Storage availability	Decentralized *monitoring* components & data-centric metrics;
Data incidents	Incident *tracing & adaptation* mechanisms for data handling modules;
Different data stores	On-demand data *integration* & access load balancing *optimization*.

Table 2 Application context and impact on architectural design

Application context	**Impact on architectural design**
Data analytics mode (batch, stream)	Dynamic storage configuration policies, plugins for application & edge-cloud data connectors;
Fault-tolerance level	*Replication* mechanism & *re-routing* strategy;
Data-sensitive applications	*Secure & verifiable* support for data storage/exchange;
Multi-model data types	Runtime customization of data *operations & processing utilities*.

One of the challenges coming from proposed engineering principles is that results for edge analytics tasks (e.g., regarding time and space complexity) might not be as expected, since all principles cannot be implemented and checked within the current project. In this context, however, proposed approaches can help all interested researchers and industry to improve revealed dependencies within edge data services or take it as a basis for further research and implementation.

3 Zusammenfassung Planaktualisierung

Alle Anpassungen des Planungsdokuments kurz zusammengefasst

The following adjustments have been made in the planning document:

- Switched position of activities planned between time periods 2019Q3 and 2019Q2.
    - After successfully finishing research activities (2018Q4 and 2019Q1) that led to the entire EDMframe architecture, using the research findings in the technique selection for edge data recovery, we realized the importance of continuing with activities that should support elastic edge data services within the EDMframe. Thus, ongoing activities (2019Q2) include further analysis for arhitecturing elastic edge storage services.
- The activities (for 2019Q3), postponed due to reasons above, might include a new approach for making an accurate near real-time decisions with incomplete data evaluating the three cases for making decisions, namely with raw data, based on representative data sample and with already planned approximate data representation.