netidee STIPENDIEN

Data Management Strategies for Near Real-Time Edge Analytics

Endbericht | Call 13 | Stipendium ID 3793

Lizenz CC-BY

Inhalt

1 Einleitung

The rising demands for the Internet of Things (IoT) sensors and devices are already an integrated part of our lives through various applications such as smart homes, smart manufacturing, healthcare, and smart cities. These IoT systems and applications often require near-real-time decision-making processes. However, traditionally performing decision-making processes in distant and geographically distributed cloud data centers presents important challenges including the transfer of large amounts of sensor data over the Internet.

Edge computing has been proposed as a promising methodology and solution enabling data processing closer to data sources. Using edge nodes such as micro data centers, edge gateways, or single-board computers like Raspberry Pis, it is possible to significantly minimize costly data transmission over the network and provide needed computation while extending cloud functionalities to the edge of the network.

Future IoT systems and applications should work intuitively and make timely and proactive decisions like human bodies but based on collected sensor data and their analysis. However, performing data analysis at the edge of the network faces multiple challenges: (i) edge infrastructure still has to cope with the rapidly growing amount of sensor data while having limited storage, computation and network capabilities; (ii) IoT systems are highly distributed in the environment making data collection and their timely analysis difficult; (iii) IoT and edge systems have higher failure probabilities compared to reliable cloud data centers (e.g., node failures, network failures, sensor aging, changes in external conditions); (iv) edge nodes have to meet different strict service level objectives (SLOs) coming from modern applications (e.g., data/service availability, resilience to failures, low latency); (v) sensor data are often incomplete and performing analytics on such data can impact the quality of decision-making processes.

In this dissertation, I aim to overcome these challenges from a data-centric perspective by proposing a generic data management framework integrating prediction-based algorithms and strategies for reliable decision-making processes at the edge of the network. Driven by data trends, probability, and approximate data analytics, modern IoT systems should take proactive decisions before the system enters undesired states asking for appropriate subsequent actions. Unfortunately, the planned completion of the work in September 2020 will be difficult to achieve due to the remaining implementations and publishing activities that still have to be done. During the summer of 2019, I had a great opportunity to gain valuable experiences as a research Intern at IBM Research Ireland, validating my proposed concepts and working on self-adaptive edge-cloud analytics placements for meeting application requirements. On the other hand, also considering challenging working conditions during recent months due to the Covid-19 lockdown, the final completion of the PhD dissertation is postponed and planned for Q1/2021.

2 Allgemeines

This dissertation deals with near real-time decision-making processes in resource-constrained computing environments at the edge of the network (e.g., micro data center, Raspberry Pi). In the IoT concept, different devices equipped with sensors and actuators are connected to the Internet, but the core element in managing any IoT system is data. Managing any IoT system generally includes sensor data collection, data processing, and acting based on the obtained results. Data processing is traditionally done in massive cloud data centers. Here, I target different data management services, important for bringing novel features into the design of future edge nodes deployed near the source of data. From an architectural design viewpoint, it is necessary to identify which data should be kept at the edge, how long to store them, and which data processing utilities and novel algorithms can assist these problems in distributed and decentralized environments. Running data processing on distributed and decentralized edge nodes poses challenges to satisfy current application requirements compared to the traditional cloud environment and on-demand manual services.

Many existing data management frameworks neglect proactive system behavior that represents one of the key elements for future trends in managing IoT systems. With our proposed strategies, future IoT systems will be proactive, able to produce more accurate and timely decisions, e.g., improving air quality and energy efficiency in smart homes, or helping doctors and patients in treating different diseases while integrated into the smart healthcare systems). Although there are several studies available, the state-of-the-art still lacks efficient data management strategies for near real-time edge analytics, especially, in targeting predictive analytics as an important methodology for revolutionizing IoT systems that include critical decision-making processes.

The following **research questions** guided this work:

1. *Which specific elements of an edge data management framework are necessary to enable near real-time analytics?* Enabling near real-time analytics requires investigation of edge requirements such as latency, accuracy, resource capabilities, and consequently detecting and proposing appropriate components, modules, techniques, and algorithms.
2. *Which amount of data is necessary to perform accurate near real-time decisions while dealing with space limited storage?* Limited edge capacities represent one of the

critical bottlenecks for accurate edge analytics, requiring novel approaches to balance the quantity of data stored with the quality of near real-time decisions.

3. *How to efficiently recover incomplete datasets, while coping with near real-time requirements from IoT systems?* Missing or invalid measurements in a dataset are often and they can highly affect the quality of decisions for many IoT systems. Timely and accurate recovery of incomplete datasets at the edge is a critical task.
4. *How to manage elastic edge storage services at runtime for data-driven decision-making in edge deployed systems?* This leads to the investigation of novel approaches for elasticity management of autonomous edge storage services as well as the definition of elastic operations according to data workload characterizations and existing data management strategies.
5. *How can different data representations improve predictive analytics and decision-making processes at the edge?* The increasing amount of data and their high dimensionality in modern IoT applications make an analysis of the entire datasets difficult. The major idea is to make an approximation of the input sensor data using other representations such as symbolic data representation.
6. *How to efficiently track dataset movements in highly distributed edge environments and adaptively guide the placement of analytics tasks?* This problem can be found in typical applications such as object detection in video surveillance where the dataset of a specific camera can at different timepoints be stored in different locations (distributed edge storage nodes), preventing developers from timely and accurately executing queries or other analytic tasks.

The main goal of this work is to provide a generic data management framework with integrated strategies and methods that should enable efficient and near real-time edge data analytics. Edge analytics requires investigation of different strategies including data collection, movement, storage, analytics approximation, and data management, utilizing various prediction methods and performing accurate decision-making processes. Besides novel edge concepts and strategies showing the theoretical contribution, I aim to provide their practical applicability through novel algorithms and simulations with sensor data coming from real-world IoT systems.

3 Ergebnisse

So far, during the funding period, contributions and results of work are published in two high-quality venues, namely, IEEE Transactions on Services Computing (IEEE TSC) and European Conference on Software Architecture (ECSA). Titles and short descriptions of achieved results are as follows:

- Lujic, I., De Maio, V., & Brandic, I. (2019). Resilient Edge Data Management Framework. IEEE Transactions on Services Computing. The official article is currently available as an Early Access contribution on IEEE Xplore website. (DOI: https://doi.org/10.1109/TSC.2019.2962016).

  Here, we propose complete EDMFrame, a framework featuring a generic mechanism for recovery of multiple gaps in incomplete datasets, using single-technique recovery (STR) and multiple-technique recovery (MTR) involving projection recovery maps (PRMs). We further devise an adaptive storage management mechanism for reducing data stored at the edge, keeping only the data necessary for predictive analytics. We conduct experiments using time series from smart buildings, (i) automatically recovering various multiple gaps and reducing errors up to 65.48% with MTR compared to STR; (ii) reducing amounts of data stored to 39.9% on average, keeping prediction accuracy around 98.83%.

- Lujic, I., & Truong, H. L. (2019, September). Architecturing elastic edge storage services for data-driven decision making. In European Conference on Software Architecture (pp. 97-105). Springer, Cham. This work is published as a part of the Lecture Notes in Computer Science book series (LNCS, volume 11681). (DOI: https://doi.org/10.1007/978-3-030-29983-5_7).

  For better support of future edge analytics, in this work, we analyze requirements and dependencies in edge data services together with edge data analytics support. We propose a novel, holistic approach for architecturing elastic edge storage services, featuring three aspects, namely, (i) data and system characterization (e.g., metrics and key properties), (ii) system operations (e.g., elasticity and data quality management) and (iii) data processing utilities (e.g., data approximation and prediction). In this regard, seven principles for the architecture design and engineering of edge data services are presented.

These publications contain answers to research questions 1-4 from the previous section. Considering remaining research questions on efficient edge analytics placement, we proposed SAPLaw, a data Locality-aware approach for Self-adaptive Analytics Placement. This allows tracking datasets' movements across edge nodes while allowing dynamic deploying of data analytics tasks that depend on the dataset locations, helping developers to efficiently schedule required analytics across different infrastructures. Further, other preliminary results show promising solutions making an approximation of the input data while keeping essential features of interest for effective edge data analytics. This can directly impact the quality of decision-making processes in time-sensitive IoT systems such as smart buildings and smart healthcare.

During the funding period I regularly submitted monthly blogs about my research progress as well as new challenges, implementations and difficulties reached on the way. All blog posts can be found on my netidee-website (https://www.netidee.at/data-management-strategies-near-real-time-edge-analytics), while Table 1 shows a summarized list of blog topics during the funding period.

Table 1: Overview of the blog posts published

#	Date	Title
1	30.11.2018	Moving to the edge...
2	26.12.2018	Finding methodologies for overcoming emerging challenges
3	20.01.2019	State-of-the-art
4	18.02.2019	From raw sensor data to smart actuator decisions
5	17.03.2019	Towards Self-adaptive Technique Selection for Edge Data Recovery
6	15.04.2019	Need for Elastic Edge Data Services
7	15.05.2019	Investigating Elasticity for Edge Storage Services
8	15.06.2019	Engineering Principles for Edge Data Services
9	15.07.2019	Dealing with Approximate Data Representation and Analytics
10	15.08.2019	From the Raw Time Series to a Symbolic Data Representation
11	15.09.2019	Impact on Architectural Design of Elastic Edge Storage Services
12	15.10.2019	Intuitive and Proactive IoT Systems
13	15.11.2019	Edge Data Management Solutions and Limitations
14	15.12.2019	Practical applicability of EDM strategies
15	15.01.2020	Contributions accepted for publication in IEEE TSC
16	15.02.2020	A Testbed Setup for Edge Data Management Strategies
17	20.03.2020	Self-adaptive Analytics Placement - Use Case
18	15.04.2020	Self-adaptive Analytics Placement – Architecture Overview

4 Geplante weiterführende Aktivitäten

In the initial plan, finishing my dissertation was in September 2020. However, there are several activities and challenges that had an impact on changing the initial time plan. During the summer of 2019, I had valuable and useful experience as a research intern at IBM Research Ireland. Also considering challenging working conditions during recent months of the Covid-19 lockdown, this work is not yet completed, and consequently, writing my PhD dissertation is still waiting for more research and publishing activities to be done. Although a major part of the research is finished, there are several open activities in my schedule:

- In **Q3/2020** and **Q4/2020** I plan to focus on final experimental procedures and remaining implementation parts as well as submitting work done in two conference papers. The remaining work includes (i) validation of the symbolic data representation on edge analytics tasks; (ii) implementation of self-adaptive analytics placement based on data locality; (iii) completion of downsized, lightweight ML models for resource-constrained edge infrastructures. The order of solving these issues will be defined depending on findings and deadlines for venue submissions.

- The greatest bottleneck can be the waiting time for positive feedback from planned conference/journal publication venues. However, possible final writings and revisions of PhD dissertation are planned for **Q1/2021**. Once PhD dissertation is officially accepted, it will be available on the project website under CC-BY according to the netidee-agreement.

5 Anregungen für Weiterführung durch Dritte

This work represents an important cornerstone for future IoT systems and edge applications development, aiming to efficiently cope with (i) rapidly growing amounts of IoT sensor data; (ii) limited computation and storage capabilities of edge infrastructures; and (iii) strict requirements from IoT system to perform near real-time decision-making processes. The promising results show important insights for **system integrators** and **developers** that are responsible for combining different system subcomponents into the edge-cloud pipelines and creating dynamic adaptations and actions for efficient edge data processing.

Proposed mechanisms in this work bring high scientific values as well as practical values when dealing with real-world problems, but also open a space for new theoretical aspects important for the edge communication infrastructure. For example, considering contributions on the adaptive recovery of incomplete sensor data coming from real-world smart buildings, the proposed algorithms demonstrate valuable solutions and can directly impact all smart building similar IoT systems in the **industrial sector**. Besides theoretical concepts that are available to the **scientific community** through several published conference and journal papers, the most important algorithms are also accessible through GitHub open source.

IoT solution architects can also utilize given insights from this work and directly improve decision-making processes in IoT systems considering data movement, connectivity, communication, data management, and storage. Consequently, this can lead to having more intuitive systems and making the necessary transformation from reactive to proactive IoT systems as well as initiating the creation of novel applications through different IoT research and development (**R&D**) departments.