Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

Big Data Inspired Water Management Platform for Sensor Data

Viktor Dmitriyev[1], Jorge Marx Gómez[2] and Manuel Osmers[3]

Abstract: The usage of the data generated by the various types of sensors is addressed to be one of the promising source of improvements and one of the core stones of the Industry 4.0 initiative of the German government. The domain of sensor data management continuously demonstrates it's ability to improve currently existing processes in various directions. Despite theoretically applicability of the same approach of the sensor data management in the domains other that water, current work focuses on demonstrating preliminary results of the research done with focus on water management. The preliminary results demonstrated in the form of principal platform architecture. The main idea behind the paper is to show the principal architecture in order to open a future opportunity for building a prototypical implementation of the proposed architecture later on. The principal architecture has an big data phenomena principals like volume, velocity and variety by-design.

Keywords: sensors, smart data, big data, water management, internet of things

1 Introduction

Water resource is a starting point of live for all living species on our planet. In case of the space missions in searching extraterrestrial life, the "target number one" is searching for availability of water. Most life on earth need freshwater, but only 0.3 % of the water is on the planet calculated as drinkable [Gl93]. So it is important to use our water sustainable and sensible. But despite there are a lot of discussions about the water all around the globe, the main question that still needs an answer is "How we can support water sustainable usage with information and communications technologies (ICT)?" Despite the topic of resources sustainability is not a new one, the water as most important resource for human, nature and industry needs more attention from different viewpoints, an example of such is ICT. Water is no longer an endless resource; rather we have to think about dependable methods and look for approached to enhance sustainability aspects of water utilization.

It is a common and widespread idea among people living in Germany that they have no problems with the water supply, but within the next few years people with high probability may face some of the problems such as other countries are facing.

[1] University of Oldenburg, Business Information Systems / VLBA, Ammerländer Heerstr. 114-118, 26129 Oldenburg, viktor.dmitriyev@uni-oldenburg.de
[2] University of Oldenburg, Business Information Systems / VLBA, Ammerländer Heerstr. 114-118, 26129 Oldenburg, jorge.marx.gomez@uni-oldenburg.de
[3] University of Oldenburg, Business Information Systems / VLBA, Ammerländer Heerstr. 114-118, 26129 Oldenburg, manuel.osmers@uni-oldenburg.de

Especially, the nitrate concentration in the water of the North West region in Germany is a promising problem, next to salinization and population growing, which will defiantly attracting far more attention in the future. Some institutions are working on the topic of prevention such problems. They started to use information and communications technologies to visualize and compute the scenarios, in which they simulate the problems that are about to appear. By conducting different simulations and visualizations they are trying searching for applicable solutions.

The other side of the promising ICT utilization within the context resource management, and water management in particular, is the usage of other data sources like Open Data, as well as services and interfaces developed on top of the Open Data [Da10, MRS13, Gu11]. The Open Data is emerging as one of the most promising value adding data-centric approaches. Open Data is attracting more and more attention from the business and government sides. In 2011 McKinsey predicted that big data market for the European Sector by 2020 is going to be around €250 billion per year [Ma11], where the Open Data is playing a really great role. The further steps for the work are to combine Open Data with proprietary data from water suppliers and state authorities in order to create more sustainable and value adding processes heavily supported by decisions made upon data.

Especially, the concepts of openness within the Open Data approach is promising to make cities ”smarter”. An idea of making cities smarter attracts more attention to the ”Smart Cities”[To13] phenomena. Following ideas and concepts behind Smart Cities allows resource such as urban city services, their infrastructure utilization to being more sustainable, which is very important, even crucial, with the continuously growing urbanization [Ki14, Ha12]. Despite usage of the water is very common in almost any locations, the possibility of proper and most optimal water resource utilization without much investments (with just ICT driver innovations) is very promising in highly populated and well organized locations such as cities.

2 Main Idea

The usage of the data generated by the various types of sensors, which can be a part of the day-to-day used smartphone device, or some other dedicated devices that is able to measure temperature, or gas pressure within the car engine, are agreed to be one of the promising source of improvements and one of the core stones in the Internet of Things (IoT) [AIM10, Gu13] and Industry 4.0 initiative of the German Government [Ka13]. The domain of sensor data management continuously demonstrates it’s ability to improve the current existing processes in various directions.The current work is focused on demonstrating preliminary results of the research done with focus on water management. The idea of the work is first of all to show the principal architecture and than build a prototypical implementation of the proposed architecture later on. The principal architecture has an big data principals like volume, velocity and variety. Current work is going to demonstrate the principal architecture that will enlarge the potential of water management, in particular, the architecture is addressing following requirements (a) consolidate sensor and event data, (b) store data, (c) offer interactive analysis and standard reporting, (d) offer pluggable components approach.

In the water management there are number of different measured rates that can be traced by particular sensor (e.g. pressure, fluid-flow, level of water pollution, salt water intrusion, etc.). By adding on top of the sensor data, a business data from the companies and the data from customers a very interesting scenarios for the water management can be create by the experts from various domains (e.g. sustainability). Such huge amounts on data from many different sources creates a 3 mostly common dimension of the big data: Volume, Velocity and Variety [La01]. On the other hand, the approach is designed for usage of business and customer data for economic and sustainable goals as well, for example by incorporating them together into prediction models or sustainable reporting. Due to the above mentioned the work has a relevance to the Environmental Management Information System (EMIS). An implementation of the use cases with EMIS inside proposed architecture should demonstrate the applicability of ideas described by the current work on more granular level.

The idea is to demonstrate abilities of ICT to help in solving many types of problems, which water management is currently facing and going to face in the future. Taking into consideration addressed architecture will help the organization to be more flexible in terms of data storage and processing, as well as prepare the infrastructure, which is mainly focused on the data analysis part, for future changes and generation of enormous amounts of sensors data. The general approach proposed by the work to meet addressed requirements can be divided into following steps: (a) consolidate historical water related data; (b) consolidate real time water data from sensors; (c) visualize water reservoirs with respect to their parameters volume, geo-location, water consumption and water recovery potential; (d) apply statistical methods (e.g., data mining, machine learning, etc.).

3 Architecture Overview

General overview of the architecture of the Water Management Platform for Sensor Data is shown on the Fig.1. The core component of the overall architecture is an In-Memory Database, which can be defined as an database management system (DBMS) that utilizes capacity of main memory and uses disks as the secondary storage [PZ12]. Usage of main memory as the primary data storage allows to overcome physical obstacle of data storages based on the magnetic rotating disks. The usage of In-Memory Database will significantly simplifications data model (because most of the works around such as pre-aggregations within the analytical cubes are not more required), it also should be bring a significant performance boots to the all types of queries (reporting, add-hoc, OLAP, etc.).

The primary roles of the Water Management Platform for Sensor Data is to offer centralized data storage, integration and analysis facilities for each particular water related scenario. Part of the platform, that is responsible for the data storage consists out of 5 main components and additional ability of plugging custom component: (1) Event Processing Engine; (2) Long-Term Storage; (3) In-Memory Database; (4) Reporting; (5) Interactive

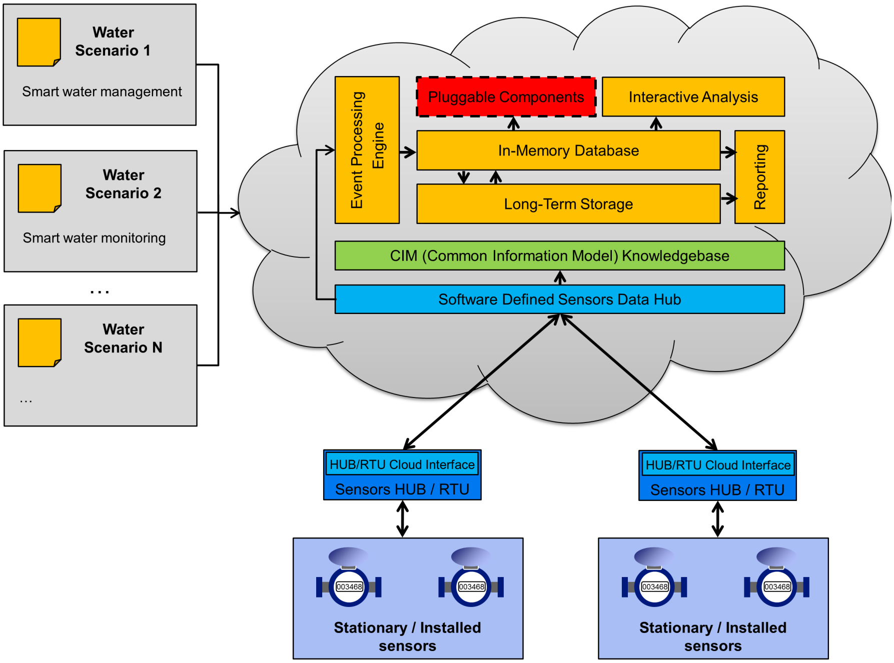

Fig. 1: Architecture Of Water Management Platform for Sensor Data

Analysis. The platform gets sensor data through the Software Driven Sensor Data Hub. The internal communications between components within the platform are facilitated by commonly acceptable standards and protocols (OData, ODBC, HTTP/S, etc.). In the particular case, communications between Software Driven Sensor Data Hub and Sensors/RTU-s will be based on CIM (Common Information Model) that brings interoperability to the target architecture and also support computability with a wider range of devices.

As it's shown on the Fig.1 software components are run inside the private cloud, such approach allows to scale the overall architecture with accordance of demand. Using enhanced cloud solution stack will also prepare a future landscape for the deployment of the targeted solution. Integration of advanced sensors, such as RTUs (Remote Terminal Unit), in the architecture through the Sensor Data Hub, as it's shown in Fig. 1, will bring flexibility to the architecture in the sense of integrating new types of RTUs (existing on the marked and future ones).

As it was mentioned above, every component of the platform should run in the cloud stack. The Event Processing Engine can be implemented with the help of framework that allows processing of continuous data, for example Odysseus Framework [Ap12]. The Long-Term Storage cane be implemented as a distributed database or as a distributed file system with processing engine on top. In particular, the role of the Long-Term Storage can be overtaken by e.g. Apache Hadoop and other components from it's Ecosystem [Va13]. An In-Memory Database will be used for the use case, in which requirements are demanding data processing heavy operations such as ad-hoc queries, KPI visualization or even data mining and machine learning. Interactive Analysis of data will integrate differed analytical capabilities such as highly customizable analytical views. The predictive analytics can be as well integrated into the platform through Plugable Component. Reporting can be done by means of any openly available tool such, for example Pentaho BI [LBF14] or even some other in-house developed software solution.

4 Related Works

There are number of projects contributing to the sustainability of the particular city and making a city ”a smart” one by integrating various of data streams within a single system and processing integrated data in order to achieve better level of understating of ongoing events and processes within the particular environment and context. One of the such projects is the CityPulse [Ko14, Ga15], the project that was funded by the EU and was conducted by the multinational consortium. The major goal of the CityPulse project is to benefit from interconnected ecosystem of the city and improve sustainability by more efficient (optimal) operations performed with city services and infrastructure. CityPulse tried [Ko14, Ga15] to integrate various source of the streaming data generated by the devices together with social data streams for in a about city into an single large scale real-time system.

The other example of the related works is the project called CyberWater [CMI13]. The main goal of the project is to conduct a system, which is able provide a decision support in risk related environments focusing on an integrated management of watersheds and pollution caused by some accidents. Due to the overwhelming amounts of data and geographical heterogeneity of the data sources (sensors), the core system is seen to be a distributed and based on the heavily based on the Service-Oriented Architecture (SOA) principles for interoperability.

Despite there are a list of similarities between the proposed Water Management Platform for Sensor Data and CyberWater in main goals, there are a list of principal differences between them on architecture level. The main goal of the CyberWater is to help in decisions, whether the main goal of the proposed approach is to simplify working with data on all levels and give a strong and robust tool to the domain experts to work with data without investing much time into establishing a proper data engineering pipeline. Another difference to be mentioned is an interaction between components in proposed approach. A particular interaction is done through previously defined protocols, which works with significant efficiency in comparison with Service-Oriented Architecture (SOA), because the decision whether to use or not to use particular protocol for connections, as well as depth of integration into protocol of an component, can be done based on various requirements (performance, fault-tolerance. etc.). Such elasticity in protocols integration make proposed architecture flexible. Other significant difference is that the Water Management Platform for Sensor Data is user case independent. One of the main focus of the proposed approach is to make a transfer process of every single data point, which is generate by sensors, transparent for the end users.

5 Conclusions and Future Works

Current work demonstrates demand for the platform, which is not only able to manage water related data, but is also able to face upcoming challenges addressed by fast evolving initiatives such as Industry 4.0 and Internet of Things. The general overview of the principal architecture for the Water Management Platform for Sensor Data was shown and described. Future work will be the detailed description of the water use cases and the prototypical implementation of the proposed platform together with an industry partner, which are showing a great interest in conducting a prototype based on the proposed architecture and foreseeing a huge benefit for themselves and their customers.

References

[AIM10] Atzori, Luigi; Iera, Antonio; Morabito, Giacomo: The internet of things: A survey. Computer networks, 54(15):2787–2805, 2010.

[Ap12] Appelrath, H.-Jürgen; Geesen, Dennis; Grawunder, Marco; Michelsen, Timo; Nicklas, Daniela: Odysseus: A Highly Customizable Framework for Creating Efficient Event Stream Management Systems. In: Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems. DEBS '12, ACM, New York, NY, USA, pp. 367–368, 2012.

[CMI13] Ciolofan, Sorin N; Mocanu, Mariana; Ionita, Anca: Distributed Cyberinfrastructure for Decision Support in Risk Related Environments. In: Parallel and Distributed Computing (ISPDC), 2013 IEEE 12th International Symposium on. IEEE, pp. 109–115, 2013.

[Da10] Davies, Tim: Open data, democracy and public sector reform. A look at open government data use from data. gov. uk, 2010.

[Ga15] Ganz, Frieder; Puschmann, Daniel; Barnaghi, Payam; Carrez, Francois: A Practical Evaluation of Information Processing and Abstraction Techniques for the Internet of Things. 2015.

[Gl93] Gleick, Peter H et al.: Water in crisis: a guide to the world's fresh water resources. Oxford University Press, Inc., 1993.

[Gu11] Gurstein, Michael B: Open data: Empowering the empowered or effective data use for everyone? First Monday, 16(2), 2011.

[Gu13] Gubbi, Jayavardhana; Buyya, Rajkumar; Marusic, Slaven; Palaniswami, Marimuthu: Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7):1645–1660, 2013.

[Ha12] Hancke, Gerhard P; Hancke Jr, Gerhard P et al.: The role of advanced sensing in smart cities. Sensors, 13(1):393–425, 2012.

[Ka13] Kagermann, Henning; Helbig, Johannes; Hellinger, Ariane; Wahlster, Wolfgang: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group. Forschungsunion, 2013.

[Ki14] Kitchin, Rob: The real-time city? Big data and smart urbanism. GeoJournal, 79(1):1–14, 2014.

[Ko14] Kolozali, Sefki; Bermudez-Edo, Maria; Puschmann, Daniel; Ganz, Frieder; Barnaghi, Payam: A Knowledge-based Approach for Real-Time IoT Data Stream Annotation and Processing. In: Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE. IEEE, pp. 215–222, 2014.

[La01] Laney, Douglas: 3-D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, February, 6, 2001.

[LBF14] Lapa, Joaquim; Bernardino, Jorge; Figueiredo, Ana: A Comparative Analysis of Open Source Business Intelligence Platforms. In: Proceedings of the International Conference on Information Systems and Design of Communication. ISDOC '14, ACM, New York, NY, USA, pp. 86–92, 2014.

[Ma11] Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Byers, Angela Hung; Institute, McKinsey Global: Big data: The next frontier for innovation, competition, and productivity. 2011.

[MRS13] Magalhaes, Gustavo; Roseira, Catarina; Strover, Sharon: Open Government Data Intermediaries: A Terminology Framework. In: Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance. ICEGOV '13, ACM, New York, NY, USA, pp. 330–333, 2013.

[PZ12] Plattner, Hasso; Zeier, Alexander: In-Memory data management: Technology and applications. Springer Science & Business Media, 2012.

[To13] Townsend, Anthony M: Smart cities: big data, civic hackers, and the quest for a new utopia. WW Norton & Company, 2013.

[Va13] Vavilapalli, Vinod Kumar; Murthy, Arun C; Douglas, Chris; Agarwal, Sharad; Konar, Mahadev; Evans, Robert; Graves, Thomas; Lowe, Jason; Shah, Hitesh; Seth, Siddharth et al.: Apache hadoop yarn: Yet another resource negotiator. In: Proceedings of the 4th annual Symposium on Cloud Computing. ACM, p. 5, 2013.