

Proposal of a Big data System for an Intelligent Management of Water Resources

Sara Bouziane^{1*}, Badraddine Aghoutane¹, Aniss Moumen², Ali Sahlaoui³, and Anas EL Ouali⁴

¹Informatics and Applications Laboratory, Science Faculty of Meknes, Moulay Ismail University, Meknes, Morocco

²Ibn Tofail University, National School of Applied Sciences, Kenitra, Morocco

³Laboratory of Geo-Engineering and Environment, Faculty of Sciences, Moulay Ismail University, Meknes, Morocco

⁴Department of Environment, Functional ecology and environmental engineering laboratory, Faculty of Sciences and Technology, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Abstract. Today, advanced technologies like Big Data, IoT, and Cloud Computing can provide new opportunities and applications in all sectors. In the water sector, water scarcity has become a common concern of different institutions and actors worldwide. In this context, several approaches and systems have been proposed and developed, using these technologies, allowing intelligent water resources management. Internet of Things can be used for assisting the Water Industry to collect data, manage and monitor the water infrastructures using smart devices. Big Data is a strategic technology for analyzing and interpreting collected data into valuable and helpful information for better decision making. This paper presents Big Data and Internet of Things technologies. It addresses their uses in some use cases such as municipal water losses, water pollution in agriculture, water Leak detection, etc., to provide new systems and innovative solutions for intelligent water resources management. Based on this study, we propose a Big Data and IoT architecture for intelligent water resources management.

1 Introduction

The scarcity of water is a real challenge for the whole world. With population growth, climate change, increased demand for food and energy, establishing a water resources management plan has become a global priority in the 21st century.

Since independence, water has been considered an essential component of the economy in Morocco. According to The North Africa Post, on January 14, 2020, King Mohammed VI launched a program of 115.4 billion dirhams to secure the water supply until 2027 [1]. "This program aims to diversify water resources to meet demand, guarantee water security and combat the effects of climate change", has mentioned in the same post [1].

Water data are of many types and can be collected from many sources. It concerns surface water, groundwater, and the climate.

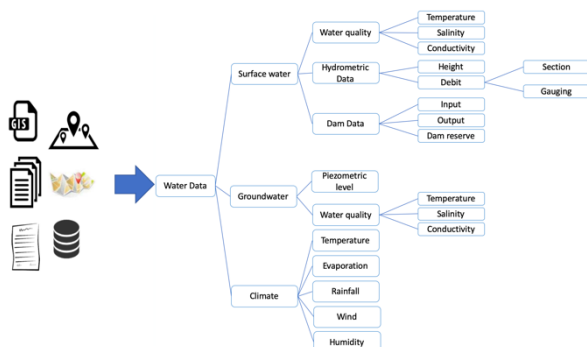


Fig. 1. Water Data [3].

Fig.1 above shows the different types of data relating to water resources [2].

In recent years, Big Data has aroused great interest in public debate and the social sciences. It designates unconventional strategies and innovative technologies used for capturing, managing, processing, and analyzing a massive volume of data [2].

Indeed, analytics and Big Data could be decisive in our fight against the loss of water. Combined with the Internet of Things (IoT), they could help us optimize the consumption of resources, reduce their losses.

This paper discusses, first of all, the use of technologies like Cloud Computing, IoT, and Big Data by water actors in Morocco, then, it presents an overview about Big Data and two use cases in water resources management:

- Municipal water losses.
- Water pollution in agriculture.

Afterward, it presents IoT technology and its applications in the water sector.

Finally, this paper details the steps to work with Big Data and proposes an architecture of a Big Data system for intelligent water resources management.

2 The use of Cloud Computing, IoT, and Big Data by water actors in Morocco

* Corresponding author: sa.bouziane@edu.umi.ac.ma

A study entitled "The Use of Cloud Computing, IoT, and Big Data by Water Sector Actors in Morocco: Between Myth and Reality" was carried out and communicated by Pr. A. Moumen in the second edition of the International Congress of I SEE Geomatics [2].

Water stakeholders in Morocco were invited to give their opinion on IoT, Big Data, and Cloud Computing.

This study showed that stakeholders believe that technologies like Big Data, IoT, and Cloud Computing can positively impact several use cases for intelligent water resources management, but they are not being used.

The next step in this study was to organize working sessions with water stakeholders to explain to them the contribution of these technologies in the intelligent management of water resources and how to deploy such solutions.

3 Big Data - Overview

Data has an intrinsic property; it grows at ever-increasing speeds. Nowadays, we are talking about a large volume of data of different formats and types.

The term Big Data first appeared in the early 90s. Its presence and importance are increasing exponentially over time. It designates unconventional strategies and innovative technologies used for capturing, managing, processing, and analyzing a vast volume of data [2].

3.1 Types of Big Data

In Big Data, there are three types of data; structured, semi-structured, and unstructured data.

3.1.1 Structured data

Structured data conforms to a tabular model with a relationship between its rows and columns, making it easier to analyze [2].

Primary examples of this type of data are Excel files or SQL databases.

3.1.2 Semi-structured data

Data that does not conform to a data model but has organizational properties such as tags and other markers that make it easier to analyze, aka XML or JSON [2].

3.1.3 Unstructured data

Unstructured data is data of an unknown form that cannot be analyzed or stored in RDBMS unless it is transformed into a structured format [2].

3.2 Growth of data

Data sources have evolved from websites to Cloud Computing and IoT while passing through social networks, smart devices, and online services.

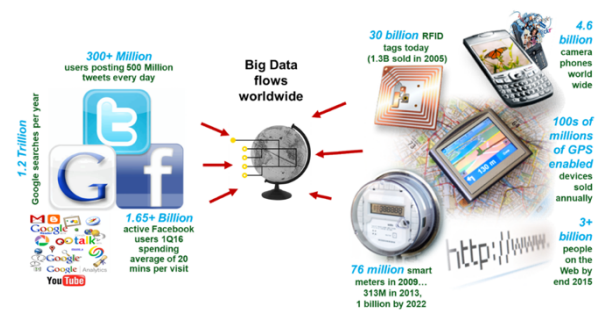


Fig. 2. A large volume of data from different sources.

This evolution has caused massive growth of data at increasing speeds. Nowadays, we are witnessing a data tsunami.

4 Big Data use cases: Water resources management

By 2050, scientists predict, the demand for food will double, and the world's population will reach 9 billion people.

Water scarcity is a significant obstacle to increasing food production, especially with droughts and other climate change effects that make water scarcer.

Because of that, water resources managers have started to look for water resources management strategies using last technologies like Big Data and IoT.

Indeed, Analytics and Big Data could prove decisive in our fight against the loss of water.

4.1 Municipal water losses

Water leaks due to the obsolescence of some municipal piping networks constitute a significant problem causing waste of water resources.

Water stakeholders can use Big Data systems for intelligent control of pipeline networks. They can be informed in real-time about the state of the city's pipes and pipelines.

Using specialized sensors, agents can read the state of the infrastructure and locate the source of the water loss and, therefore, intervene on these leak points.

We can even anticipate and prevent these leaks and water losses by using Machine Learning techniques.

4.2 Water Pollution in Agriculture

The use of pesticides (based on chemical components) in agriculture is one of the leading causes of water pollution.

To deal with this problem, we can minimize the number of chemical inputs used in agriculture and limit the discharge of pollutants into the water.

Machine Learning makes it possible to accurately assess each plant's specific needs to fight against pests that threaten plant growth.

To collect this information relating to plants, in real-time, we can use drones equipped with hyperspectral cameras capable of flying over agricultural fields.

We can transmit and process this data by Machine Learning algorithms to be synthesized to help farmers to spread fertilizers and pesticides optimally according to the needs of the identified plants.

5 IoT applications in water management

The proliferation of wireless technology and the increasing efficiency of small, embedded systems led to a technological convergence known as the Internet of Things (IoT) [4]. Atzori et al. note in 2010 that the primary strength of the IoT concept is the impact it will have on different aspects of daily life and the behaviour of potential users [5].

5.1 IoT Solutions

We can define an IoT solution as a set of devices and sensors connected to a cloud platform via a gateway [2]. This cloud platform allows us to manage, store, secure, and analyze the large volume of data to extract useful information and insights from it [2].

In this paper, we use the term 'IoT solution' to describe a fully working setup for IoT.

Fig. 3 represents the general architecture of an IoT solution. It includes four interdependent layers: (1) IoT Devices, (2) Communication Protocols, (3) Data Processing, and (4) IoT Applications (Fig. 3).

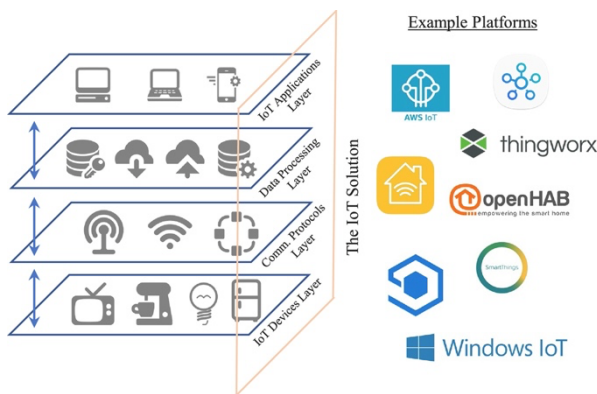


Fig. 3. The essential four layers within an IoT solution in different IoT platforms.

IoT devices are equipped with embedded sensors, actuators, processors, and transceivers that permit whole interaction with the physical world [4].

Connectivity is the essential element of an IoT solution. In order to guarantee the reliability of the IoT communications, many protocols can be used within the same IoT implementation to accommodate the environment's constraints (e.g., Bluetooth, WIFI, Zigbee, Z-Wave, etc.) [4].

The Data Processing layer is an intermediary between the hardware and the Application layer. Its main functions are collecting data from devices over several protocols and network topologies, remoting device configuration and control, device management, and over-the-air firmware updates [2].

IoT applications are the user interface to communicate with the IoT solution. They allow users to read real-time data captured by IoT devices, set control values, and monitor devices.

5.2 IoT Applications in water management

Intelligent water management systems, based on the combination of IoT, Big Data, and Cloud Computing technologies, can help water stakeholders tackle water scarcity at a lower cost [6, 7].

Below are some use cases of IoT applications in water resources management.

5.2.1 Water Leak detection

Water supply is done through pipelines. Or these pipes can be prone to obsolescence, for many reasons, it can cause water leakage and therefore water wastage [8].

IoT technologies provide intelligent ways to detect leaks more precisely and increase the rate at which leaks are being detected [2].

5.2.2 Water Quality monitoring

One of the most crucial water management aspects is water quality. We can use IoT devices such as intelligent water monitoring equipment. These devices can be installed to collect and keep track of pH, turbidity, pressure, flow, and water temperature and send the data to the water management network for real-time water quality analysis [2,9].

5.2.3 Real-time Water Control

IoT technology allows water stakeholders to access, control and configure different water management operations aspects in real-time [2].

The objective is to notify the user of the water quality parameters in real-time to analyze and make the right decisions [10].

6 Steps for working with Big Data

There are five fundamental steps to complete a Big Data Analytics project [2].

6.1 Data collection

Data are collected from various sources in many ways. IoT has made real-time data collection even more comfortable and manageable [11].

IoT platforms provide devices to collect data in real-time and send it securely to the data processing layer, so it can be exploited [2].

6.2 Data streaming

Big Data streaming is the continuity of transferring large volumes of data at high speed from a source system to a target [2].

As an extension of the core Spark API, Spark Streaming provides scalability, high throughput, and fault tolerance stream in the processing of live data streams. [12].

As shown in Fig. 4, data can be collected from several sources like Kafka or Flume and can be treated using complex algorithms written with high-level functions. Processed data can be transferred to databases, filesystems, and live dashboards [12] (Fig. 4).

While streaming, data can be filtered, transformed, validated, cleansed, enriched, and even analyzed.



Fig. 4. Big Data streaming architecture.

6.3 Data storage

The large amounts of collected data must be stored securely and integrated effectively [11]. In Big Data, traditional methods (RDBMS) are not suitable for data storage [2]. Thanks to the object-based storage architecture, data collections can be stored and managed as objects; this is a flexible solution for storing semi-structured and unstructured data [13]. Cloud storage is the preferred solution for many companies; it is a highly flexible and resource-efficient technique for storing data [11]. Additionally, cloud storage facilitates data sharing [13].

6.4 Data visualization

Visualization is an essential step in the processing data lifecycle. It allows clear transmission and communication of information by graphical means [13].

Statements, charts, diagrams, graphs, and virtual reality are the most popular techniques to visualize data [14].

6.5 Data analysis

The analysis is the most crucial step in Big Data analytics. It allows end-users, by using different analysis algorithms, to discover helpful information that can help organizations and actors make the right business decisions [15].

Python is the standard programming language used by data scientists to investigate Big Data, thanks to its many useful tools and libraries, like Pandas and Matplotlib.

7 Proposed architecture

Concerning the problem described above, water scarcity, and based on IoT, Big Data, and Cloud Computing technologies, we propose a system that

water stakeholders can use for intelligent management of water resources, which the architecture is shown in Fig. 5.

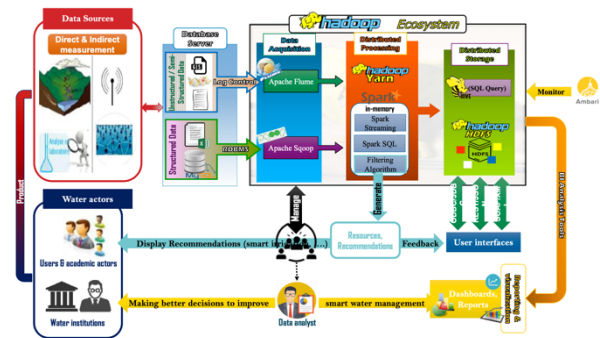


Fig. 5. Big Data system for water resources management.

Data collection can be assured from many sources with direct and indirect (IoT devices) measurements. Structured data are acquired via Apache Sqoop, while Apache Flume acquires unstructured and semi-structured data.

Sqoop is an application used to transfer data from relational databases to Hadoop.

Flume is a secure, distributed, and accessible service that allows us to efficiently collect, aggregate, and transfer large volumes of continuous event data.

Apache Hadoop is an open-source framework used to facilitate interaction with big data.

Spark is used for data streaming to generate resources and recommendations and display them on the user interface.

Apache Hive is used for Data storage. It is the data warehouse of Hadoop for providing data queries and analysis.

The Hadoop ecosystem allows us to generate graphs and analyze to make the right decisions for intelligent water resources management.

8 Conclusion

In this paper, we have focused on the common concerns to water scarcity by different institutions and actors. We have presented an overview of advanced technologies such as Big Data, IoT, and Cloud Computing and how they can be used for water resources management.

We have proposed an architecture of a Big Data system for water management that will be adapted in consultation with the partners and validated once we have all the required information.

References

1. North Africa post <https://northafricapost.com/36925-king-mohammed-vi-launches-115-4-billion-dirham-program-to-secure-water-supply-up-to-2027.html#A->

2. A. Moumen, et al., *L'usage du IoT, Cloud Computing et Big Data spatiale chez les acteurs du secteur de l'eau au Maroc : entre mythe et réalité*, in the second edition of the International Congress of ISEE Geomatics, 25-27 March 2019, Palace of Sciences, Monastir – Tunisia (2019)
3. 3ème Edition des Journées Doctorales en Systèmes d'Information, Réseaux et Télécommunication (JDSIRT), ENSIAS, +- Université Mohammed V, 12-13 Novembre 2015, Rabat, Maroc (2015)
4. L. Babun et al., *Comput. Netw.*, Volume 192, **108040** (2021)
5. L. Atzori, A. Iera, G. Morabito, *Comput. Netw.*, pp. 2787-2805 (2010)
6. <https://www.digiteum.com/smart-water-management-iot/#>
7. M. Singh and S. Ahmed, IoT based smart water management systems: A systematic review, *Materials Today: Proceedings* (2020), <https://doi.org/10.1016/j.matpr.2020.08.588>
8. P. Swetha Reddy et al., *Water leakage detection monitoring and controlling system using IoT*, *IJET* 7(2.7), pp. 120-123 (2018)
9. K. Spandana1, V.R. Seshagiri, *Internet of things (IoT) based smart water quality monitoring system*, *IJET*, 7(3.6), pp. 259 – 262 (2018)
10. A. S. Punpale and P. B. Borole, *Water quality monitoring and control using IoT and industrial automation*, *IJSTE*, 4(12), pp. 133 – 238 (2018)
11. M. Azeem, A. Haleem, S. Bahl, et al., Big data applications to take up major challenges across manufacturing industries: A brief review, *Materials Today: Proceedings*, <https://doi.org/10.1016/j.matpr.2021.02.147>
12. Spark Apache website <https://spark.apache.org/docs/latest/streaming-programming-guide.html>
13. F. Tao et al., *JMS* **48**, pp. 157–169 (2018)
14. Mittal et al., *Proc IMechE Part B: J Engineering Manufacture*, 233(5), pp. 1342–1361 (2017)
15. S. Sossi Alaoui, et al., *Air pollution prediction through the Internet of Things technology and Big Data Analytics*, *IJCISTUDIES*, Vol. 8, **3** (2019)