Big Data Allowing next Generation Fish Farming Systems

Mohamed El Mehdi El Aissi, Sarah Benjelloun, Yassine Loukili, Younes Lakhrissi and Safae Elhaj Ben Ali
SIGER Laboratory, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Keywords: Big Data, Data Lake, Data Management, Data Insights, Smart Fish Farming, Data Processing.

Abstract: To tackle the increasing needs of fish farming production, it becomes a necessity to adopt new techniques of digitalization in order to handle the huge quantity of data generated by the fish farming ecosystem. This step will allow fish farmers and researchers to extract valuable information and as a result they will improve their productivity. Although Big Data techniques are affording many advantages, it has not been widely applied in agriculture, especially in fish farming. This article sheds light on the potential that Big Data have in the fish farming industry. It shows the functional and technical need of including Big Data in traditional fish farming systems. Thereafter, it presents the architecture proposal of a smart fish farming system based on Big Data technologies.

1 INTRODUCTION

Population growth, socioeconomic factors and food industry are highly correlated. According to Sarker et al., the world population has grown from three to six billion, which created a high food demand (Sarker, Md Nazirul Islam, et al. 2020). The Food and Agriculture Organization of United Nations (FAO-UN) has affirmed that population is growing by 30% until 2050, consequently food production must be increased by 70% (United Nations 2015).

Along with this massive increase there are many challenges and constraints such as land degradation, water contamination and the degradation of fishery resources, which creates real uncertainties of food security (Sarker et al. 2020) (United Nations 2015).

To Handle these increasing demands, many researcher papers have been published since the 1990s. However, most of the studies and initiatives were focused on developing the agricultural field only. Nowadays, researchers are trying to merge between agriculture and technology such as Remote Sensing, Big Data, Cloud Computing and the Internet of Things (IoT). Indeed, all these initiatives are converging to the concept of "Smart Farming".

Big Data Analysis has been used in various industries such as banking, insurance, medicine, industry and marketing. Despite all the success that Big Data has achieved in the mentioned fields, it started being applied to agriculture only recently, but not in the fish farming domain. According to some agriculture specialists and corporations, applying Big Data to fish farming could increase the profit and production quality (Lytos et al. 2020).

For this main cause, adopting new Big Data technics has become a necessity in order to tackle the challenges of productivity, environmental impact, food security and sustainability (Sarker et al. 2020) (Lytos et al. 2020).

The motivation for writing this article stems from the fact that Big Data is a modern technique that is not treated yet in fish farming despite its benefits as it proved a significant potential in other domains.

Indeed, the main goal of this contribution is to present a focused overview of the existing challenges of fish farming and how Big Data can be used to overcome those challenges by offering a dedicated Big Data Architecture for the fish farming systems.

2 RELATED WORKS

The existing research works are mostly focused on agriculture and how can we extract valuable information from collected data. In their paper, Lytos et al. considers agriculture as a complicated scientific field by its nature that requires a dedicated framework for managing the incoming data and processing it.

Moreover, they presented a comparison between many frameworks used to handle the agricultural data

El Aissi, M., Benjelloun, S., Loukili, Y., Lakhrissi, Y. and Elhaj Ben Ali, S.
Big Data Allowing next Generation Fish Farming Systems.
DOI: 10.5220/0010739000003101
In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, pages 518-522
ISBN: 978-989-758-559-3

but, among 14 frameworks only two are supporting Big Data. Indeed, a dedicated Big Data architecture is required as it allows the integration of IoT, and this will allow an efficient data collection through sensors and the automation of many operations (Lytos et al. 2020).

In parallel, N.N.Misra et al. has showed in their article that since we integrated IoT solutions in agriculture it starts generating massive amount of data, that is categorized as “Big Data", and this may open new opportunities to monitor agriculture and food process. They presented how we can merge between Big Data, IoT and AI to shape the future of agri-food systems. In general, this study has focused on agriculture analysis, mainly, it exposes how greenhouses can be monitored using AI and how Supply Chain modernization can enhance the food quality (Misra et al. 2020).

Even though they did not present how Big Data techniques participates in developing the fish farming field through IoT integration.

Furthermore, in his research paper Xinting Yang et al. has presented how Deep Learning (DL) can be applied for smart fish farming to handle the series of challenges for data processing. In addition, he affirmed that the most significant contribution of DL is its ability to automatically extract features (Yang et al. 2021). Although, before we start using DL on this data, we should be able to manage it, especially when we talk about fish farming massive data that is generated by IoT devices. To overcome this situation, creating a dedicated Data Lake architecture is the best solution to manage data efficiently, in order to be consumed in a better way.

3 BIG DATA TECHNIQUES FOR FISH FARMING

In general, we can differentiate between two concepts of database management systems.

The first concept is the data warehouse, which is simply a relational database designed especially for querying and analyzing data. Usually, it contains structured historical data coming from transactional databases; furthermore, all data stored is structured. Additionally, the organization of data in a data warehouse is subject oriented, which means each table is built to respond exactly to a need that was already specified. One more thing to remember is that before storing the data in the staging zone of a DWH it should respect an exact structure and may pass through data cleansing steps to ensure high data quality in reporting (M.M.El Aissi et al. 2020).

In smart fish farming, data and information are the core elements. The aggregation and advanced analytics of all or part of the data will lead to the ability to make scientifically based decisions. However, the massive amount of data in smart fish farming imposes a variety of challenges, such as multiple sources, multiple formats and complex data (Fleming, Aysha, et al. 2018).

Multiple sources include information regarding the equipment, the fish, the environment, the breeding process and people. The multiple formats include text, image and audio. The data complexities stem from different cultured species, modes and stages. Addressing the above high-dimensional, nonlinear and massive data is an extremely challenging task. Moreover, a data warehouse has a special data collection approach known as ETL that has three main steps: Extraction, Transformation and Loading.

Data extraction refers to extracting data from homogeneous/heterogeneous sources; data transformation involves processing data in order to cleansing it and transforming it to an adequate structure, so that querying, and analysis are simple and efficient; finally, data loading means the ingestion of data into the final target.

The second concept is the data lake; it is defined as a powerful Big Data architecture for storing huge amounts of structured, semi-structured and unstructured data. Furthermore, it allows storing every data type in its original format regardless of its size and its source (data as-is storing) (Kour et al. 2020).

Data Lake allows a high flexibility as the data structure or schema is not defined when data is captured. This means we can store all the data without a careful design or the need to know what the future use case is and where our collected data should provide answers (Panwar et al. 2020). It must be noted that this architecture will allow a high suppleness while interacting with data like SQL queries, Big Data analytics, full text search, real-time analytics, and machine learning.

In a data lake, we are talking about an ELT (Extraction, Transformation and Loading) process rather than an ETL (Extraction, Loading and Transformation), as presented for the data warehouses, which is completely evident since the most important phase in a data lake architecture is to gather data and store it as much as we can (M.M.El Aissi et al. 2020), then when a use case appear and depending on the use case we can move to the

transformation phase where we start building the models by combining the different existing data.

The table below provides a quick view on the main differences between the Data warehouse and the Data Lake architecture:

Table 1: Data Warehouse versus Data Lake.

Characteristics	Data Warehouse	Data Lake
Data	Relational	Non-Relational and Relational
Schema	On-Write	On-Read
Price	Higher cost	Lower cost
Data Quality	Clean data	Raw data
Users	Data Analyst	Data Scientist / Data Analyst / Data Developer
Analytics	Batch reporting, BI and Data visualization	Machine Learning, Predictive Analysis, Data Discovery

In order to choose the adequate architecture, we must highlight that the collected fish farming data has some characteristics which differentiate it from the data that was used to be stored in the traditional database management systems, since it is massive data.

Besides, there is a long debate on the data which is generated from fish farms. Some researchers do not consider it as Big Data because of its characteristics. They pointed out that fish farm data does not fit properly with the existing characteristics of Big Data. But other researchers consider it as Big Data and point out that it will be Big Data when all data from the fish farming system are pooled together (Sarker et al. 2020) (Panwar et al. 2020). Considering both views, this part will explain the characteristics of fish farm data as follows.

- **Volume:** Generally, fish data is generated from various equipment from tanks, pumps and manual measurement and sensors (Hassan et al. 2020). It is usually stored in office computers or even cloud services. Statistics show that huge data are generated by fish farming systems and stored according to specific years. So, the volume of data is so big and even sometimes difficult to move to other devices.
- **Velocity:** Velocity means the rapidly changing characteristics of data. Generally, 10 MB of data are generated from multiple sensors per hour. The size of data keeps increasing depending on fish production activities (Sagar et al. 2018). So, the data size is increasing at the pick of fish production but continues even in low production. Furthermore, the duration or lifetime of fish is varying from one to another. So, some fish have short lifetime, but some others have long lifetime. Consequently, fish data vary from each other.
- **Variety:** Variety means a range of data sources. Fish farms data should be categorized according to its source and this will help to understand and use the data efficiently. Sometimes, data varies from equipment to equipment, automated sensor to manual method (Sagar et al. 2018) (Hajjaji et al. 2021). So, data should be managed properly according to its collection, types, and procedures. Manual data should be transferred properly to the computer and integrated with machine-based sensors (Lioutas et al. 2019). The manually collected data should be analysed and transformed into structured format for future use.
- **Veracity:** In most cases, fish farming data are unstructured in nature. Manual data is more unstructured than other data which is generated by sensors (Lioutas et al. 2019). There is also a problem about the data quality. Usually manually collected data are messier than machine-based sensor data (Lee, Junchan, et al. 2020). Furthermore, fish farming data are greatly influenced by sensor factors and human factors and it can be minimized by taking proper records of manually applied inputs and installing automated sensors with monitoring carefully.
- **Value:** Usually, fish farming data bears a high volume of information which is generated by every stage (Schuster, Jason. 2017). Moreover, it has great potential and value for making future decisions (Carbonell, Isabelle. 2016). The data sources are usually from sensors, APIs, post-production studies and environmental factors (Pham et al. 2018). It possesses great value in different stages in fish production like water pH, water nutrient content, feeds content, humidity, temperature requirement, diseases and other related valuable information (Majumdar et al. 2017).

The above statements clearly explain that fish farming data is really Big Data in all aspects of Big Data characteristics, so adopting the Data Lake architecture is a necessity. The potential of fish farming Big Data is actually dependent on its use by agriculturists, fish farmers, researchers, academicians, and decision makers.

4 PROPOSED FISH FARMING DATA LAKE ARCHITECTURE

To illustrate the need of Big Data in a fish farming system, we propose a technical architecture that can handle the data flows generated in usual fish farming operations. In this architecture we distinguish three main phases (Figure 1):

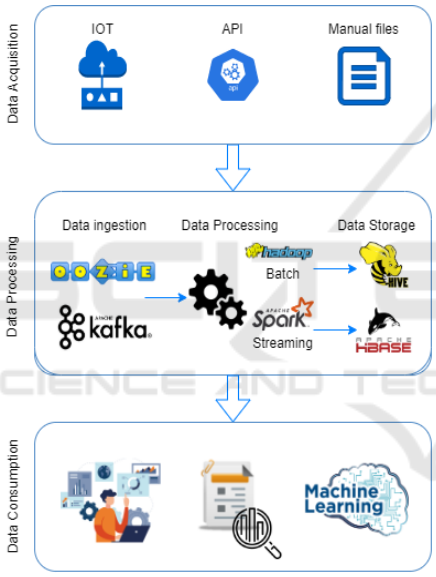

Figure 1: Data management process.

4.1 Data Acquisition

Technical architecture is a minimal set of rules and interactions of the parts or elements in order to ensure that the system satisfies a specified objective as well as a set of requirements. For this purpose, we proposed a technical architecture corresponding to the Big Data system for smart fish farming.

Fish farming architecture generates three types of data. The huge amount of data that sensors produce must be processed before it can be used. Since there are multiple devices, data need to be transformed or standardized to a unified format. The sensors communicate and send data through HTTP/HTTPS or MQTT protocols to the cloud. The data regards mainly: temperature, light, chemical composition and average weight of fish (Majumdar et al. 2017).

Flat files contain complementary data manually written in csv format and sent through email. The files are received and sent to the data lake cluster via mail bot also called data bot. Manual data can be food quantities per tank or number of fish.

APIs data consist of all the external data regarding the weather, fish prices, fish food prices and others. It is collected from APIs using curl shell command and stored into flat files in order to be ingested.

4.2 Data Ingestion / Processing

Before Storing, analyzing and accessing data, data ingestion gathers data and brings it to the processing system.

Two tools can be distinguished based on the type of data:

- Apache Oozie is used to orchestrate a process. It is used to build and schedule data jobs at a precise date and time or frequency. Flat files from emails and APIs are handled in long running jobs in order to store them to the Hadoop cluster (Pham et al. 2018). These jobs are scheduled daily to ensure that new updated data is used in dashboards and reports.
- Apache Kafka is responsible for consuming and fetching the streaming data produced by sensors and published to the OPC server (Pham et al. 2018).

Batch data from APIs and manual data fetched by Oozie is pushed to the data lake into Hive tables. Apache Spark is then used to access, transform and store data into Apache HBase tables.

4.3 Data Consumption

The Data can be consumed in many types as Data Visualization through dashboards presenting KPIs for each uses case (Zhou, Chao, et al. 2019) (Zhao, Jian, et al. 2018), we can perform some predictive analysis through Machine learning or even some statistical analysis (Salman, Ahmad, et al. 2020).

5 CONCLUSION AND PERSPECTIVES

The integration of technology in fish farming is considered as a key element for maximization of fish farming production in order to support the high food

demand of the world population that is keep growing each year. Despite the huge benefits of Big Data application in many fields including agriculture, it is still not very accessible in fish farming activities. Big Data techniques are the pilar for transforming traditional fish farming to modern digital fish farming. It overcomes all the limitations related to fish farming systems by analysing farmer's need, market need, financial efficiency and other stakeholder perspectives. The study is highlighting the necessity of integrating Big Data in fish farming then presenting a dedicated Data Lake architecture for fish farming use case. Besides, strong initiatives are a necessity to tackle the related challenges such as data quality, data availability and data governance. Our future works are focused on using the proposed data lake architecture as a base for an advanced study using different types of data analysis including artificial intelligence and machine learning.

REFERENCES

Carbonell, I. (2016). The ethics of Big Data in big agriculture. *Internet Policy Review*, *5*(1).

Fleming, A., Jakku, E., Lim-Camacho, L., Taylor, B., & Thorburn, P. (2018). Is Big Data for big farming or for everyone? Perceptions in the Australian grains industry. Agronomy for Sustainable Development, 38(3), 1-10.

Hajjaji, Y., Boulila, W., Farah, I. R., Romdhani, I., & Hussain, A. (2021). Big Data and IoT-based applications in smart environments: A systematic review. *Computer Science Review*, *39*, 100318.

Hasan, M. (2020). Real-time and low-cost IoT based farming using raspberry Pi. *Indonesian Journal of Electrical Engineering and Computer Science*, *17*(1), 197-204.

Kour, V. P., & Arora, S. (2020). Recent Developments of the Internet of Things in Agriculture: A Survey. IEEE Access, 8, 129924-129957.

Lee, J., Angani, A., Thalluri, T., & jae Shin, K. (2020, January). Realization of Water Process Control for Smart Fish Farm. In *2020 International Conference on Electronics, Information, and Communication (ICEIC)* (pp. 1-5). IEEE.

Lioutas, E. D., Charatsari, C., La Rocca, G., & De Rosa, M. (2019). Key questions on the use of Big Data in farming: An activity theory approach. *NJAS-Wageningen Journal of Life Sciences*, *90*, 100297.

Lytos, A., Lagkas, T., Sarigiannidis, P., Zervakis, M., & Livanos, G. (2020). Towards smart farming: Systems, frameworks and exploitation of multiple sources. Computer Networks, 172, 107147.

Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of Big Data. *Journal of Big Data*, *4*(1), 1-15.

Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2020). IoT, Big Data and artificial intelligence in agriculture and food industry. IEEE Internet of Things Journal.

Mohamed El Mehdi El Aissi, et al. "Data Lake versus Data Warehouse: A comparative study." Accepted on WITS 2020:https://www.springer.com/gp/book/9789813368927

Panwar, A., & Bhatnagar, V. (2020). Data lake architecture: a new repository for data engineer. International Journal of Organizational and Collective Intelligence (IJOCI), 10(1), 63-75.

Pham, X., & Stack, M. (2018). How data analytics is transforming agriculture. *Business horizons*, *61*(1), 125-133.

Sagar, B. M., & Cauvery, N. K. (2018). Agriculture data analytics in crop yield estimation: a critical review. *Indonesian Journal of Electrical Engineering and Computer Science*, *12*(3), 1087-1093.

Salman, A., Siddiqui, S. A., Shafait, F., Mian, A., Shortis, M. R., Khurshid, K., ... & Schwanecke, U. (2020). Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. *ICES Journal of Marine Science*, *77*(4), 1295-1307.

Sarker, M. N. I., Islam, M. S., Murmu, H., & Rozario, E. (2020). Role of Big Data on digital farming. Int J Sci Technol Res, 9(4), 1222-1225.

Schuster, J. (2017). Big Data ethics and the digital age of agriculture. *Resource Magazine*, *24*(1), 20-21.

United Nations, ―World Population Prospects -Population Division -United Nations, World Population Prospects. pp. 1–5, 2015.

Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S., & Zhou, C. (2021). Deep learning for smart fish farming: applications, opportunities and challenges. Reviews in Aquaculture, 13(1), 66-90.

Zhao, J., Li, Y., Zhang, F., Zhu, S., Liu, Y., Lu, H., & Ye, Z. (2018). Semi-supervised learning-based live fish identification in aquaculture using modified deep convolutional generative adversarial networks. *Transactions of the ASABE*, *61*(2), 699-710.

Zhou, C., Xu, D., Chen, L., Zhang, S., Sun, C., Yang, X., & Wang, Y. (2019). Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. *Aquaculture*, *507*, 457-465.