An Architecture of a Data Lake for the Sharing, Agricultural Knowledge in Burkina Faso

Safiatou Sore[1](✉), Yaya Traore[2], Moustapha Bikienga[3], and Frédéric T. Ouedraogo[3]

[1] Ecole Normale Supérieure, Koudougou, Burkina Faso
sore_safiatou@yahoo.fr
[2] Université Joseph KI-ZERBO/IBAM, Ouagadougou, Burkina Faso
[3] Université Norbert ZONGO, Koudougou, Burkina Faso

Abstract. With the advent of big data today, agricultural research institutes have stored enough data in databases at every level. This data contains a lot of knowledge that is often not well known by researchers and on the other hand, it may contain hidden semantic relationships. Therefore, it is important to find an adequate storage system for these data in order to discover new useful information that can be shared between the different actors in the field to improve agricultural productivity in Burkina. The concept that best responds to this storage and analysis problem is the data lake. Indeed, data lakes offer the possibility of storing a large volume of data in any format and data structure. They also provide services for data access and analysis. Our work revolves around the integration and storage of agricultural data (structured, semi-structured and unstructured). With the plurality of agricultural storage systems, the heterogeneity of the actors in the field and the disparity of the tools leads us to think that the data lake is a solution.

The purpose of this essay is to investigate data lakes and their requirements while also putting out a broad framework for an agricultural data lake in Burkina Faso.

Keywords: agriculture · data lakes · architecture · ingestion · exploration · treatment · questioning · metadata

1 Introduction

Agriculture in developing countries is characterized by low harvests and, consequently, some limited yields [1].

The main activity of the population in Burkina Faso is agriculture. It employs 86% of the active population. However, traditional agricultural techniques and climate conditions are still present.[2]. Variations in temperature and meteorology have impacts on all fields and in particular on agriculture in general. Indeed, in the absence of adequate rainfall and appropriate temperatures, harvests are poor. To better boost agriculture in Burkina and increase its yield, a redefinition is necessary.

In order to improve agricultural productivity and yield in Burkina Faso, the Institut de Environnement et de Recherches Agricoles du Burkina Faso (INERA) is one of the

R. A. Saeed et al. (Eds.): AFRICOMM 2022, LNICST 499, pp. 209–218, 2023.
https://doi.org/10.1007/978-3-031-34896-9_13

four specialized institutes of the Centre National de la Recherche Scientifique et Technologique (CNRST) in charge of agricultural and environmental studies and research. INERA's missions are to generate knowledge and technological innovations to improve plant, animal, forestry, wildlife, and fisheries production; to develop and conduct research for the sustainable management of natural resources; to provide technical support for agricultural, environmental, and forestry development; and to link research and development, including technology transfer and innovation. Teams of researchers are constantly collecting data for their studies. In addition, it should be noted that technical agents, depending on their specialities, collect and disseminate farm data, meteorological data, soil data, environmental data, etc.... Due to the diversity of this data, a global situational analysis seems difficult to achieve.

The digitization of agriculture, which has emerged since the mid-2010s, is the answer to this need. It defines agriculture and, more broadly, a food system through the use of digital sciences and technologies such as data acquisition (satellites, sensors, connected objects, smartphones, etc.), transfer and storage (3G/4G/5G coverage, low-speed terrestrial or satellite networks, clouds), and on-board or remote processing (supercomputers accessible via very high-speed communication networks, artificial intelligence) at all levels of agriculture[3].

In fact, the introduction of new tools and technologies, combined with agronomic, meteorological and environmental knowledge, could help improve productivity. This is why, taking into account the extension and dissemination of this data, it becomes essential to cope with the adversity of nature and improve agricultural yields. It is necessary to use the explicit knowledge of researchers, engineers, and technicians who have various results in order to make them accessible and, at the same time, contribute to improving agriculture in Burkina Faso. To do so, we propose the following steps: (i) opening up agricultural, meteorological and environmental research data; (ii) linking these data by properly structuring domain and process knowledge; and (iii) finally, scientific knowledge of climatic, agronomic and territorial data will be combined and shared among the different actors.

The approach we propose for the sharing, extension and dissemination of agricultural knowledge in Burkina Faso is a data lake approach.

The data lake domain (used as a staging area or data source for data warehouses or used as an experimental platform for scientists or data analysts) is booming and providing a solution in terms of storing large quantities of structured, unstructured or semi-structured data with the possibility of storing different types of data in their original format [4]. In addition, this type of platform poses no constraints in terms of file size or category. It enables high-performance data analysis and native integration.

In agriculture, Madera and Al. [5] try to answer the following questions about data lakes.

What is this entry-level concept? What is its definition? How does it compare to traditional data warehousing and analysis architectures? What are the main components of this architecture and how do they impact the design of agricultural information systems in general?

Following this introduction, Sect. 2 below provides the background. Section 3 then presents a list of data lake definitions, a review of existing data lake architectures and in

Sect. 4, general architecture suggestions for building an agricultural data lake. Finally in Sect. 5, the conclusion of the article followed by some perspectives.

2 Context

The development and use of agricultural information requires the mobilization of extensive sets of knowledge. Computer or agricultural knowledge, contextual or general, theoretical or empirical, explicit or tacit, etc. all this knowledge that the expert has and that the novice lacks.

Taken in isolation, the information contained in the processing descriptions is not knowledge. They only become so when they are linked to other knowledge that makes it possible to decide on the action to be taken to achieve the considered need [11]. Thus, a novice user who would only have a simple catalog of processing descriptions could not make the chain of decisions that leads to the realization of his need. In this context, three objectives emerge: the collection of agricultural data which are massive and highly heterogeneous for better monitoring of agricultural production; the enrichment of this data because of its diversity, with each other to bring out hitherto unknown information which will be used to better understand agricultural knowledge such as: discovering the cause of performance of certain areas over others; the spatial distribution and concentration of crops according to their yields and cultivation periods; data analysis and visualization to provide specialists with biological data.

A computer system is therefore essential to support the analysis of such a quantity of data and several constraints will have to be taken into account by offering: a unified solution allowing the storage and analysis of the different documents at the same time; a wide range of advanced analyses, allowing in particular to summarize, describe and compare groups of documents. However, it should be noted that the analyses will be carried out by business users, it is essential to offer them an autonomous solution and therefore not requiring any particular skills in data processing.

Finally, the desired tool for the storage and analysis must be extensible so as to support the later on axes and types of analyses not yet determined.

On the other hand, the concept able to best respond to the mentioned constraints is the data lake. The big evolution of data lakes is that it makes data processing more operational, as it is able to react to data in real time. In addition, it should be noted that the data lake manages both structured and unstructured data, and offers a very wide range of analyses.

3 Data Lake Architecture Review and Requirements

Today, the concept of data lakes is used in academic and personal contexts. In the area of infrastructure management, they are starting to grow [6]. They are also recognized as a new direction in agricultural data management.

The concept of a data lake is quite new and was introduced by Dixon in 2010 who uses the following analogy as a definition: “If you think of a data mart as a bottled water store - cleaned and packaged and structured for easy consumption - the data lake is a great quantity of water in a more natural state. The contents of the data lake flow from a source to fill the lake, and various users of the lake can come and examine, dive or take samples" [7]. This concept provides an answer to the management of Big Data as a data store that can store indefinitely very large amounts of raw data in their original format. This storage methodology facilitates the coexistence between different schemas and data structures. Generally, this storage concerns objects or blobs (Binary Large Object), which is a mass of data in binary form that does not necessarily conform to a file format. All company data is stored in a single data lake. In addition to transformed data, there is raw data, which contains copies of data from various source systems. This data is used for reporting, visualization, analysis or machine learning.[8]. The data lake diverges from data warehousing, which usually translates into an integrated, highly structured, subject-oriented BI database. In addition, the data warehouse has the shortcoming of dividing data into hermetically sealed silos [9].

However, over time, various definitions and synonyms of this concept have been discovered in the literature, as data hub [10], data reservoir [11], data laboratory [5].

Finally Hail et al. [12] in their proposal define a data lake as: "a flexible and scalable data storage and management system that ingests and stores raw data from heterogeneous sources in their native format, and provides on-the-fly query processing and data analysis. On the basis of this definition, we retain that the data lake: (i) stores raw data. (ii) is not just a simple storage system; (iii) supports on-demand data processing and querying; (iv) a human expert must be able to provide additional data about the data source if the metadata cannot be automatically extracted from its source.

A data lake should retain [13] the raw formats of the data and provided: services that make it possible to integrate any type of data; cataloguing services to make the data accessible according to well-defined metadata standards; data management metadata according to well-defined standards; data management; logical and physical design and scalability.

With respect to data lake architecture, existing research describes the structure and components of the system that specify the storage, organization, and use of data. [14–16] report the distinction between basin or zone or layer architectures. A recent survey of data lake architecture by Sawadogo et al. [16] confirmed this.

The basin architecture [17] distributes the ingested data according to its state and use. The ingested data is first stored in a raw data pool that processes it. The data pool is actually a staging area, as the data is then packaged and transferred to one of the analog, application or textual data pools. The associated processes prepare the data for further analytical processing. Afterwards, the data that is not/no longer actively used, but could be useful again, is stored safely for a long time in the archive data pool.

In contrast, zone architecture [11, 18] separates the life cycle of each value set is organized in various phases. So, one could decide on the existence of specific areas for data loading and quality control, areas for storage of these raw data, cleaned and validated data storage, data discovery and exploration, or use data for business analysis/research. In the literature in general, the zone architecture considers three (03) [19] or four (04) [20] or even six (06) zones [21] and in this architecture, the components depend on the level of maturity of the data, while they follow the data format (structured data, textual data, sensor data, etc.) in the basin architecture [15, 22].

The zone architecture or the data pond architecture usually lacks technical details about the functionality, which prevents modular and reproducible implementations.

To overcome this lack, a proposal to classify the architectures into three (03) groups has been made [16]: functional, based on data maturity and hybrid. For [16] the basin architecture can indeed be considered as a variant of the zone architecture, because the location of the data also depends on its level of refinement. Following their ideas, functional architectures define the components of the data lake with respect to a set of basic functions. The basic functions are: data ingestion consisting of the transfer of data into the lake, data storage to hold raw and refined data, data processing and its accessibility.

The architectures-based on data maturity allow to plan and organize the data life cycle.

In a hybrid data lake architecture, components rely on both traditional lake functions and the degree of data refinement.

On the other hand, some authors, such as Giebler [23], believe that data ingestion, storage, and organization are not enough for a data lake, and propose a complete nine-dimensional data lake platform.

4 Proposal for a General Agricultural Data Lake Architecture

The role of agriculture is essentially to provide food for humanity. To reach these objectives, it must work with agronomy. The latter brings to agricultural activities the necessary scientific and technical knowledge.

Indeed, agronomic sciences cover agricultural production systems, including soil management, Agri-environment, plant and animal production, agricultural management and financing, as well as food processing. They seek to address the challenges of managing agriculture, territories and resources in a sustainable development perspective. Agricultural researchers in general and those of Burkina Faso in particular can be in charge of fundamental or applied research such as laboratory or station experimentation, or even the implementation of studies, surveys and the dissemination of knowledge. As such, they collect a lot of data. The data concern in particular surveys or observations in the field which are generally composed of temporal, spatial and thematic information.

In some cases, information systems are designed and developed by different research teams to facilitate the exploitation of data collected in various research programs, usually constituting autonomous and heterogeneous data sources.

In many other cases, research activities are conducted and scientific data, although informative, are transcribed into text documents or tabular files or images and videos are taken with various cameras that appear several times. These do not allow for an overall analysis of the situation, which does not facilitate the exploitation of the data.

Given the multitudes of storage sources, data sources, and varieties of data (structured, semi-structured, and unstructured), building a single storage repository seems appropriate.

Thus, to support agricultural researchers in their data management activities in an effective and efficient manner, we propose the use of a data lake, a concept introduced by Dixon in 2010 as a solution to the problems of data wake and variety.

The data lake we propose will label the data as it is ingested into the lake. Furthermore, these data are categorized according to their format and maturity level, among others. Subsequently, the user can then, on demand, group the lake data into pools or zones using the categorizations that are stored in the metadata system. For these reasons, we prefer not to base ourselves on an existing architecture, but rather to list the components that we consider important for the proper functioning of our lake.

The proposal made is the construction of an agricultural data lake that takes into account four components: (i) the ingestion of data whose objective is to facilitate the creation of metadata, (ii) the storage of data that can contain any type of data, (iii) the system of metadata allowing to find one's way in the lake, and (iv) finally the last component which is the exploitation of the data lake within which all types of use and analysis of data can be envisaged (Fig. 1).

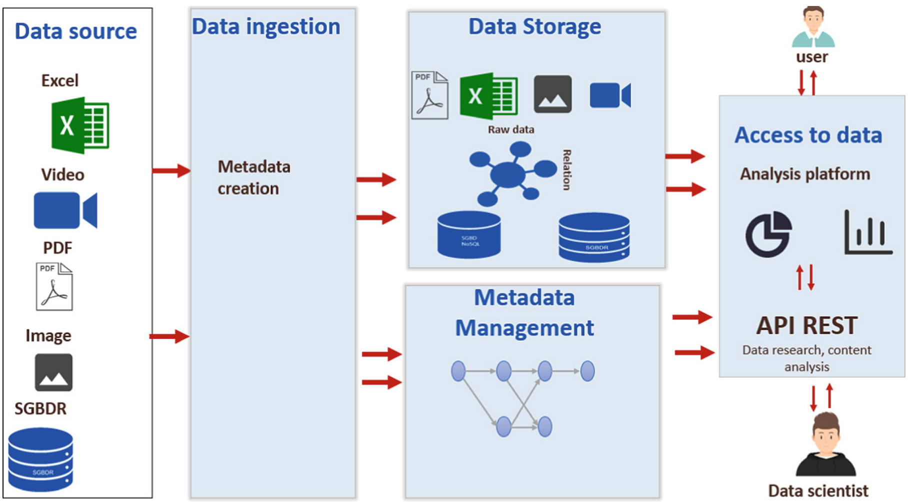

Fig. 1. Proposed agriculture data lake architecture

4.1 Data Ingestion Layer

In the literature there are three ways of transferring data: batch, real-time processing or a combination of both. In our work we will focus on batch ingestion. This method processes data based on ingestion properties. Small batches of data are merged and optimized for fast query results. Batch ingestion includes activating a pipeline to enable binding between a source and multiple storage destinations. Tools such as Sqoop, Flume (specialized in log collection, processing and transfer) and Kafka can be used for this processing [25]. However, to ensure data reliability, pipeline monitoring must be performed on an ongoing basis.

The different sources of agronomic data are very heterogeneous, especially in terms of format, type (CSV, XSLX, JPEG, HTML, RDF, etc.), naming, coding and vocabulary used. Faced with this heterogeneity, we are going to create a dictionary to list the available metadata, the redundancies, the contradictions or the complementary.

4.2 Data Storage Layer

Data storage and centralization are key objectives of data lakes and must be achieved in an agile, secure and streamlined way. All data is stored in any format (structured, semi-structured and unstructured data).

There is a wide range of technologies and techniques used and usable for storing data in lakes: Hadoop Distributed File System (HDFS) whose data is very often stored in blocks of 64 MB by default in data lakes using a file system; Relational DBMS like MySQL, PostgreSQL and Oracle are used for storing small volume structured data; Non-relational or NoSQL (Not Only SQL) DBMS (Database Management System) is also a data storage system providing scaling, distributing and replicating data across multiple machines.

The storage layer in our work is to store raw data, as well as metadata (including refined data). To do this, it exploits a combination of storage systems. Each of these systems is adapted to a specific storage. Thus, a simple file system is used to store the raw data. A graph-oriented DBMS to materialize the relationships between the metadata contained in the lake. A document-oriented DBMS to store the refined representations of the textual documents. And finally, a relational DBMS to store the refined representations of tabular and image documents.

4.3 Metadata Management

The result of a data lake should be useful knowledge for the users. It is important to prevent a data lake from becoming a data swamp, hence the importance of data governance. Metadata management is a defined process for managing data. An organization uses it to ensure the quality of that data and its availability throughout its life cycle [26]. The implementation of the data lake is the done through data governance. It provides benefits such as: (i) identifying the owner of a dataset for example and possibly having an idea of the person contacted in case of questions that directly target them; (ii) managing standardized definitions allowing a user to have an idea of the correct values of a given element; (iii) define how data is used in a business process; (iv) improve data security by protecting it with access rights; (vi) improve data traceability; and (vii) understand the life cycle of data stored in the data lake, including metadata management.

4.4 Data Exploitation Layer

Finally, the data exploitation layer enables querying of data stored in the data lake. It consists of two complementary interfaces. On the one hand, a REST API allows to analyze or extract data from the lake using a precise formalism. On the other hand, a graphical interface allows to perform the same treatments through a web application. These two tools allow a differentiated access to the lake data according to the users' profile. Thus, data scientists access the lake data through the REST API and business users through the graphical interface. This gives our data lake an inclusive character, unlike most data lakes that are data scientist oriented [26–28].

The dual interface for accessing our data takes the form of a web application, for example a client-server application accessible via a web browser. It follows a back-end-front-end architecture. This structure offers a modularity that facilitates the scalability and maintenance of the system. It also allows us to provide differentiated access to the lake data, depending on the user's skills.

The back-end interacts directly with the storage layer, from which it extracts and makes available data through microservices. These micro-services are predefined (but customizable and extensible) functions that facilitate and standardize the interrogation of the lake's data and metadata.

The front-end is in practice a graphical interface that provides intuitive and visual access to the lake data. It reuses the micro-services developed in the back-end.

The dual data access interface allows users to extract data from the lake in its raw form, or to perform a set of analyses using predefined but extensible functions, presented through services. These services can be grouped into two types: data retrieval services, on the one hand, and aggregation or content analysis services, on the other.

5 Conclusion and Perspectives

Under the pressure of the environmental crisis, seasonal risks according to the emergence and progression of crop diseases and pests, and in the face of social and democratic demands, agricultural practices and policies must evolve in Burkina Faso. Faced with this challenge, researchers in various agricultural fields and technicians are producing voluminous data (e.g., collected experiences) in different formats to stimulate this agricultural production.

The need for agricultural researchers is to have a centralized software platform capable of hosting, processing and sharing these data in different formats and from different data providers.

Advances in this direction already exist, with the platform [28] that enable management of agricultural processes and provide farmer solutions that help sustain the process in terms of spatial distribution, water management, and maintenance of mechanical systems.

The platform [29] proposes an economic model by minimizing the environmental risk of total nitrogen load as a function of irrigation water demand, land availability and food supply while maximizing agricultural economic returns and environmental fluxes; and [30] manages data and integrates low-energy IoT embedded devices to turn smart agriculture into actionable knowledge.

In addition to agricultural and farm equipment data platforms, powerful decision support tools, and comprehensive agricultural databases [31], documents and tabular files are created.

However, unlike these different platforms, the uniqueness of our approach is that it accepts multiple types of raw and heterogeneous data inputs (binary image files from cameras; tabular files, reports, and ontologies enriched with their own annotation for the process of correlation and synchronous variations of certain functions or risks of severe disorders) and allows various researchers to be able to communicate their data and combine these independent data sources to discover new knowledge.

In our work, we propose an agricultural data lake following an approach whose internal structuring is made according to the following basic functions: ingestion layer, storage layer, metadata management layer, data exploitation layer.

In the proposed data lake architecture, data is extracted from the sources and passes through the ingestion layer. There, the metadata is created and stored in the metadata system, while the data is deposited in the storage layer. The user can then browse the metadata to learn more about the data in the lake and eventually select the datasets that can be used for a specific need. He can then conduct any kind of analysis on the selected data.

Future work will continue to advance our project by doing extensive research on data and metadata management to allow different actors to easily access raw data without having to browse through large and diverse data files.

References

1. Cartier, L.E., Bürge, M.: « Agriculture and artisanal gold mining in sierra leone: alternatives or complements? »J. Int. Dev. **23**(8), 1080–1099 (2011). https://doi.org/10.1002/jid.1833
2. Dipama, J.-M. P.: « Changement climatique et agriculture durable au Burkina Faso : stratégies de résilence basées sur les savoirs locaux rapport d'étude », juin 2016, Consulté le: 31 juillet 2022. [En ligne]. Disponible sur: https://idl-bnc-idrc.dspacedirect.org/handle/10625/57568
3. Bellon-Maurel, V., Huyghe, C.:« L'innovation technologique dans l'agriculture »,Geoeconomie **80**(3), 159–180 (2016)
4. « Data Lake ou Lac de Données : définition et utilisation », Formation Data Science | Data-Scientest.com, 22 février 2021. https://datascientest.com/data-lake-tout-savoir (consulté le 13 juillet 2022)
5. C. Madera, A. Laurent, T. Libourel Rouge, et A. Miralles, « How can the data lake concept influence information system design for agriculture? ». In: 11th European Conference Dedicated to the Future Use of ICT in the Agri-Food Sector, Bioresource and Biomass Sector (EFITA 2017), Montpellier, France, juill. 2017, p. 181-182. Consulté le: 21 juillet 2022. [En ligne]. Disponible sur: https://hal.archives-ouvertes.fr/hal-01847697
6. Alekseev, A., et al.: Prototype of the Russian scientific data lake. EPJ Web Conf. **251**, 02031 (2021). https://doi.org/10.1051/epjconf/202125102031
7. « Pentaho, Hadoop, and Data Lakes », James Dixon's Blog, 14 octobre 2010. https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/ (consulté le 8 juillet 2022)
8. B. L, « Data Lake : définition, avantages et inconvénients pour l'entreprise »,*LeBigData.fr*, 10 juillet 2017. https://www.lebigdata.fr/data-lake-definition (consulté le 7 juillet 2022)
9. Stein, B., Morrison, A.: « The enterprise data lake: better integration and deeper analytics », p. 9
10. E.M.T. at E.S.S. Limited, « Introduction To The Concept Of Data Lake And Its Benefits – ESDS BLOG », 6 février 2015. https://www.esds.co.in/blog/introduction-to-the-concept-of-data-lake-and-its-benefits/ (consulté le 7 juillet 2022)
11. « Governing and Managing Big Data for Analytics and Decision Makers », p. 28
12. R. Hai, C. Quix, et M. Jarke, « Data lake concept and systems: a survey ». arXiv, 17 juin 2021. Consulté le: 8 juillet 2022. [En ligne]. Disponible sur: http://arxiv.org/abs/2106.09592
13. Sarramia, D., Claude, A., Ogereau, F., Mezhoud, J., Mailhot, G.: «CEBA: a data lake for data sharing and environmental monitoring », Sensors **22**(7), 2733 (2022)

14. Hartmann, S., Küng, J., Chakravarthy, S., Anderst-Kotsis, G., Tjoa, A.M., Khalil, I. (eds.): DEXA 2019. LNCS, vol. 11706. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27615-7
15. Giebler, C., Gröger, C., Hoos, E., Schwarz, H., Mitschang, B.: Leveraging the data lake: current state and challenges. In: Ordonez, C., Song, I.-Y., Anderst-Kotsis, G., Tjoa, A.M., Khalil, I. (eds.) DaWaK 2019. LNCS, vol. 11708, pp. 179–188. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27520-4_13
16. Sawadogo, P., Darmont, J.: On data lake architectures and metadata management. J. Intell. Inf. Syst. **56**(1), 97–120 (2020). https://doi.org/10.1007/s10844-020-00608-7
17. Inmon, B.: Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump. Technics Publications (2016)
18. [« Architecting Data Lakes [Book] ». https://www.oreilly.com/library/view/architecting-data-lakes/9781492042518/ (consulté le 26 juillet 2022)
19. Hai, R., Quix, C., Kensche, D.: Nested schema mappings for integrating JSON. In: Trujillo, J.C., et al. (eds.) ER 2018. LNCS, vol. 11157, pp. 397–405. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00847-5_28
20. Giebler, C., Gröger, C., Hoos, E., Schwarz, H., Mitschang, B.: Modeling data lakes with data vault: practical experiences, assessment, and lessons learned. In: Laender, A.H.F., Pernici, B., Lim, E.-P., de Oliveira, J.P.M. (eds.) ER 2019. LNCS, vol. 11788, pp. 63–77. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33223-5_7
21. LaPlante, A.: Architecting Data Lakes: Data Management Architectures for Advanced Business Use Cases, 1st edn. O'Reilly Media, Sebastopol (2016). Consulté le: 14 juillet 2022. [En ligne]. Disponible sur: https://learning.oreilly.com/library/view/-/9781492042518/?ar
22. Advances in Databases and Information Systems. Consulté le: 14 juillet 2022. [En ligne]. Disponible sur: https://link.springer.com/book/https://doi.org/10.1007/978-3-030-28730-6
23. « The Data Lake Architecture Framework - Digitale Bibliothek - Gesellschaft für Informatik e.V. » https://dspace.gi.de/handle/20.500.12116/35802 (consulté le 14 juillet 2022)
24. Darmont, J., Favre, C., Loudcher, S., Noûs, C.: « Data lakes for digital humanities ». In" Proceedings of the 2nd International Conference on Digital Tools & Uses Congress, October 2020, pp. 1–4 (2020). https://doi.org/10.1145/3423603.3424004
25. Mendelevitch, O., Stella, C., Eadline, D.: Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale. Addison-Wesley, Boston (2017)
26. Madera, C., Laurent, A.:« The next information architecture evolution: the data lake wave ». In: Proceedings of the 8th International Conference on Management of Digital EcoSystems, New York, NY, USA, November 2016, p. 174-180 (2016). https://doi.org/10.1145/3012071.3012077
27. Fang, H.: « Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem ». In: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), June 2015, pp. 820–824. https://doi.org/10.1109/CYBER.2015.7288049
28. Khine, p.p., Wang, Z.S.:« Data lake: a new ideology in big data era », ITM Web Conf. **17**, 03025 (2018). https://doi.org/10.1051/itmconf/201817030251
29. « Satellite data-driven multi-objective simulation-optimization modeling for water-environment-agriculture nexus in an arid endorheic lake basin ».J. Hydrol. **612**, 128207 (2022). https://doi.org/10.1016/j.jhydrol.2022.128207
30. Ouafiq, E.M., Saadane, R., Chehri, A.: « Data management and integration of low power consumption embedded devices IoT for transforming smart agriculture into actionable knowledge ».Agriculture **12**3, Art. No. 3 (2022), https://doi.org/10.3390/agriculture12030329
31. « Base de données complètes sur les exploitations agricoles : manuel de référence ». https://www150.statcan.gc.ca/n1/pub/21f0005g/21f0005g2011000-fra.htm#archived (consulté le 31 juillet 2022)