Future Generation Computer Systems 179 (2026) 108324

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Leveraging big data and cloud technology for scalable and interoperable smart farming ★

Amine Roukh *, Saïd Mahmoudi

Faculty of Engineering - ILIA / Infortech, University of Mons, Mons, 7000, Belgium

ARTICLE INFO

Keywords:
Smart farming
Big data
Cloud computing
IoT
Precision agriculture
Predictive analytics
Dairy farming
Decision support systems
Interoperability

ABSTRACT

Agriculture faces escalating demands to increase food production amid shrinking arable land, resource depletion, and climate variability. Existing smart farming solutions often lack scalability, interoperability, real-time analytics, and region-specific adaptability. This paper presents *WALLeSmart*, a cloud-based smart farming platform designed to address these challenges through a scalable Lambda architecture and a modular plugin system. Hosted on a GDPR-compliant private cloud, *WALLeSmart* integrates diverse data sources (e.g., IoT sensors, satellite imagery, weather data) to deliver real-time insights and predictive analytics, achieving low-latency processing (e.g., 80 seconds for weather data streams). Key features include a one-stop shop for accessing agricultural platforms (e.g., Myawenet, MyCDL, Cerise), a consent management system for data control, a Walloon Agricultural DataHub for secure data exchange, and a personalized dashboard for farmers. The platform's unique governance model, led by farmers, ensures autonomy and transparency. Real-world case studies in Wallonia, Belgium, demonstrate its ability to process over 3 million weather measurements and 61,130 dairy farm datasets, supporting applications like *SALVE*, *W@llHerbe*, and *MyFieldBook*. *WALLeSmart*'s generalizable design enables adaptation to diverse regions, addressing ethical concerns like algorithmic bias and data ownership through transparent AI and user-centric consent mechanisms, fostering efficiency, sustainability, and profitability.

1. Introduction

The global population is projected to grow significantly over the next several decades, reaching a peak of approximately 10.3 billion people by the mid-2080s, up from 8.2 billion in 2024 [2]. This unprecedented growth poses immense challenges for agriculture, including the need to substantially increase food production, ensure sustainable livelihoods for farmers, and protect environmental health. These challenges are further exacerbated by the shrinking availability of arable land due to urbanization and environmental degradation, the depletion of critical natural resources such as freshwater [3], and the increasingly unpredictable impacts of climate change, including extreme weather events and shifting growing seasons [4]. In this context, the agricultural sector must adopt innovative solutions to enhance productivity, sustainability, and resilience.

Precision agriculture and smart farming have emerged as transformative approaches to address these challenges. Precision agriculture leverages technologies such as global navigation satellite system (GNSS), sensors, and data analytics to optimize field-level management with remarkable accuracy [5]. Building on these principles, smart farming integrates advanced technologies like the Internet of Things (IoT), drones, and artificial intelligence (AI) to create a more comprehensive and adaptive agricultural management system [6]. By incorporating real-time data, situational awareness, and location-specific insights, smart farming enables farmers to optimize resource use, predict potential issues, and improve overall productivity [7]. Moreover, these technologies promote sustainable practices, such as precise water and nutrient management, which reduce waste and minimize environmental impact [8]. Early detection of pests and diseases through real-time monitoring and predictive analytics further reduces reliance on chemical treatments, enhancing both crop health and environmental sustainability [9].

The rapid advancement of IoT and cloud computing technologies is driving a paradigm shift in agriculture. IoT-based solutions, such as livestock and soil sensors, are becoming increasingly prevalent, providing farmers with critical data to optimize their operations [10]. For instance,

★ This work substantially extends our previous conference paper [1]. Key enhancements include a Lambda architecture for real-time analytics, integration of a modular plugin system and a one-stop shop, implementation of the Walloon Agricultural DataHub and a-Box, broader data integration (including satellite imagery), and a farmer-centric governance model. See Subsection 1.2 for details.

* Corresponding author.
E-mail addresses: amine.roukh@umons.ac.be (A. Roukh), said.mahmoudi@umons.ac.be (S. Mahmoudi).

https://doi.org/10.1016/j.future.2025.108324
Received 27 February 2025; Received in revised form 9 December 2025; Accepted 13 December 2025
Available online 21 December 2025
0167-739X/

livestock sensors enable ranchers to monitor the health, location, and behavior of animals in real-time, improving animal welfare and optimizing feed and breeding programs [11]. Soil sensors offer detailed insights into topography, moisture levels, and nutrient content, enabling precise irrigation and fertilization while reducing water and fertilizer waste [12]. These sensors also provide localized weather forecasts, helping farmers make informed decisions about planting, harvesting, and risk management. Additionally, self-driving tractors and agricultural drones are revolutionizing labor-intensive tasks. Remote-controlled tractors significantly reduce labor costs [13], while drones can reduce operation costs and an increase productivity by 83.3 % in manpower for herbicide application and 41.6 % when used for fertilizing [14].

Digital technologies, including smartphone applications and cloud-based platforms, are further empowering farmers to monitor and manage their operations remotely. These tools enable real-time tracking of equipment, crops, and livestock, as well as access to actionable insights through data analytics [15]. Cloud platforms facilitate the integration of historical data with predictive analytics, allowing farmers to anticipate market trends, weather fluctuations, and potential challenges [16]. However, the widespread adoption of these technologies generates vast amounts of data. IBM estimates that the average farm can produce up to half a million data points per day [17]. While big data holds immense potential for transforming agriculture-enabling predictive analytics, proactive decision-making, and full automation of the agri-food chain-it also introduces significant challenges. Managing this data, from acquisition and processing to storage, analysis, and visualization, requires sophisticated architectural frameworks and tools [18]. Key challenges include handling large, heterogeneous datasets from diverse sources, processing data in real-time and batch modes, applying advanced analytical techniques, and presenting insights through user-friendly interfaces.

Looking ahead, the integration of IoT with Artificial Intelligence (AI) and Machine Learning (ML) holds even greater potential for transforming the agricultural landscape. AI-driven predictive models can analyze vast amounts of data collected from sensors and drones, offering advanced insights into crop growth patterns, disease risks, and optimal harvesting times [19]. Machine learning algorithms can also continuously improve these models based on new data, allowing for increasingly precise recommendations that help maximize yield and minimize resource use. Cloud computing further enhances this by offering scalable, on-demand data storage and processing power, enabling farmers of all sizes to adopt these cutting-edge technologies without needing significant in-house infrastructure [20].

The integration of AI and cloud computing into agriculture faces several challenges, including data privacy and security concerns due to the sensitive nature of agricultural data, interoperability issues arising from diverse hardware and software systems, and data quality and management hurdles that impact the reliability of AI-driven insights [21]. Additionally, connectivity and infrastructure limitations in rural areas, high costs of implementation, and a lack of technical expertise among farmers hinder widespread adoption. Ethical concerns, such as algorithmic bias and data ownership, as well as the need for scalability and adaptability across diverse farming contexts, further complicate integration [22].

To address these challenges, this paper introduces *WALLeSmart*[1], a scalable, cloud-based smart farming platform designed to help farmers, researchers, and administrators optimize agricultural operations and decision-making. The platform integrates IoT sensors, satellite imagery, weather and milk data to provide real-time insights and predictive analytics. A key feature of *WALLeSmart* is its extensible plugin system, which allows developers and agricultural professionals to create custom applications for specific farming needs, such as optimized irrigation, fertilization, and pest control. By addressing critical challenges such as data privacy, security, and interoperability, *WALLeSmart* offers a robust and user-friendly solution for modern agriculture.

1.1. Our contributions

This paper extends our previous work [1], which introduced a preliminary framework for dairy data analytics in Wallonia. The current work presents *WALLeSmart*, a comprehensive platform with advanced interoperability and governance features. The key contributions are:

- A comprehensive survey of smart farming platforms, identifying gaps in interoperability, scalability, and region-specific adaptability.
- A scalable Lambda-based architecture integrating IoT, geospatial data, and predictive analytics, with a Walloon Agricultural DataHub for secure data exchange across platforms like Myawenet and Cerise.
- A modular plugin system and one-stop shop enabling tailored decision support systems (DSSs) and seamless access to external services.
- Implementation and evaluation of *WALLeSmart* using data from 30 dairy farms and 32 weather stations in Wallonia, with a farmer-led governance model ensuring autonomy.
- Development of a user-friendly web application with a personalized dashboard, GraphQL APIs, and an a-Box for centralized document management.

1.2. Extension of prior work

Our previous work [1] proposed a basic framework for dairy data analytics in Wallonia. This paper extends it by:

- Adopting a Lambda architecture for real-time and batch processing, reducing latency to 80 seconds for weather data.
- Introducing a modular plugin system and one-stop shop for integrating platforms like Myawenet, MyCDL, and Cerise, not present in [1].
- Implementing a Walloon Agricultural DataHub and a-Box for secure data exchange and document management.
- Expanding the data ecosystem to include weather, dairy, and potential satellite data, validated with a larger dataset (30 farms, 3 million weather measurements).
- Establishing a farmer-led governance model to ensure autonomy and transparency, enhancing trust and adoption.

1.3. Paper organization

The remainder of this paper is organized as follows. Section 2 reviews existing smart farming platforms and tools. Section 3 provides a background on smart farming technologies. Section 4 presents the design and components of the *WALLeSmart* platform. Section 5 details the implementation and evaluation of the platform using real-world data. Finally, Section 6 summarizes the key findings and outlines future research directions.

2. Related work

The rapid evolution of smart farming technologies has led to the development of numerous platforms, tools, and methodologies aimed at optimizing agricultural practices. This section provides a comprehensive review of the most relevant works in the field, focusing on their contributions, limitations, and how they relate to the proposed *WALLeSmart* platform.

Cloud computing has become a cornerstone of modern smart farming, enabling scalable data storage, processing, and analysis. *FarmBeats*, developed by Microsoft, integrates IoT sensors, drones, and cloud computing to provide farmers with real-time data on soil conditions, weather, and crop health [23]. While *FarmBeats* showcases the potential of IoT in agriculture, challenges such as technology adoption barriers and the need for farmer training remain critical for widespread implementation. Similarly, [24] paper proposes an agri-cloud framework to

[1] *WALLeSmart* can be found at: https://wallesmart.be

Table 1
Comparison of Smart Farming Platforms.

Platform	Scalable	Interoperable	Extensible	Batch	Real-time	Cost-effective	Region-Specific	Governance Model
FarmBeats [23]	✓	×	×	✓	✓	◐	×	×
Climate FieldView [26]	✓	◐	×	✓	✓	×	◐	×
farmmaps [27]	✓	×	✓	✓	×	✓	✓	◐
Barto [28]	◐	✓	✓	✓	×	✓	✓	✓
CropSense [29]	◐	×	×	✓	×	✓	×	×
PLATEM [30]	×	×	×	✓	✓	✓	◐	×
DEMETER [31]	✓	✓	✓	◐	◐	×	✓	◐
DSSAT [32]	×	◐	✓	✓	×	✓	◐	×
CropWise [33]	✓	×	✓	✓	×	×	×	×
WALLeSmart (Ours)	✓	✓	✓	✓	✓	✓	✓	✓

(✓): Supported; (×): Not supported; (◐): Partially supported.

improve the cloud computing framework of agriculture data as a service (ADaaS) over the public cloud. The model provides accurate and secure data sharing to the various stakeholders, using a security attribute-based group signature (ABGS) system. The paper of [25] introduces a Farm Management System (FMS) framework for service interoperation. The FMS allows for a marketplace of services that can share data securely. A proof of concept was developed under the *SmartAgrifood Project* and tested in a greenhouse in Crete, Greece, over nine months to identify technical issues. The FMS handles various types of data, and the complexities surrounding copyright and ownership need to be addressed. Moreover, while the paper discusses the creation of a marketplace for services, the integration of various services from different providers can be complex. Ensuring seamless interoperability among diverse applications may present technical challenges that could hinder the system's usability and effectiveness.

The applications of big data in the agriculture field are numerous, we can cite for instance: precision farming, yield prediction, risk mitigation, loss reduction, supply chain management, farm-to-fork traceability, and sustainable farming [6,34–36]. Besides, agricultural big data systems can be broadly divided into three categories: (i) advanced sensor technology systems, (ii) risk management systems, and (iii) agricultural management systems [37]. Recently, initiatives are ongoing to create agricultural platforms that collect the data needed by smart farming decision support tools. They are often created by private companies or public-private partnerships [38]. Among the former, Monsanto's *Integrated Farming Systems* platform collects data on information such as soil health and pest pressures and provide them for farmers. The Climate Corporation proposes the *Climate FieldView* platform to aggregate data of different sources in one place and provides diagnostic and applicative tools to farmers [26]. However, since 2018 only four companies dominate the market: Corteva Agriscience, Syngenta Group, BASF SE, and Bayer AG due to a series of mergers and acquisitions [39]. Regards public-privates partnerships, many countries are building systems to advance the use of modern agricultural technologies. For example, *farmmaps* (previously Akkerweb) is a Dutch web-based platform that provides access to external data sources such as weather, parcel boundaries, satellite, and data from commercials farm management system. It stores geo-referenced data, including soil maps and drone imagery. It allows combining data sources and processing of data, through a set of application modules to provide farmers decision support and recommendations [27]. Another example is the platform *Barto* in Switzerland [28]. As a stock company, *Barto* brings together public and private actors to build up a smart-farming platform that also aims to digitize operational and production farms data while avoiding duplication, which speeds up on-farm processes, reduces administrative tasks. *Barto* itself is based on *365FarmNet*, a German-based farm management software provider. Nearly 45.000 farmers are already active on this platform in Germany, Poland, Bulgaria, Austria, and France.[2] The platform operates as SaaS (Software as a service) and provides solutions for managing and recording all activities on a farm. However, those platforms are proprietary, and therefore the internal system architecture is not available. In the academic sector, many kinds of researches have been conducted in the smart farming field. The authors of [30] from Spain, demonstrate the advantage of a tool, called *PLATEM*, that applies real-time decisions from data such as variable rate irrigation, and selected parameters from field and weather conditions. Data is processed in a decision-making system based on learning prediction rules. *PLATEM* is a collaborative tool where farmers can post their experiences. It can monitor the different parameters of soil and environment to better the growth of crops. Those tasks can be accomplished through a chain of web-based graphical tools. Nevertheless, data collection is not detailed and MySQL database management is used to store it. Similarly, *CropSense* uses machine learning algorithms to analyze historical and real-time data, providing recommendations for planting and harvesting [29]. Another initiative, *DEMETER*, focuses on addressing interoperability challenges in smart agriculture [31]. By integrating diverse hardware and software resources, *DEMETER* facilitates seamless data exchange and knowledge sharing across the agri-food chain. The platform has been tested in two trials in the Murcia region of Spain, demonstrating its benefits for arable crop management. Despite these advancements, many existing solutions continue to face challenges related to data heterogeneity, real-time processing, and user-friendly visualization-factors that are crucial for widespread adoption by farmers [18].

The work of [40] utilize IoT sensors to monitor electrical conductivity, pH, temperature, and humidity, enabling precise irrigation and fertilization. Similarly, studies of [41,42] employ IoT devices to track animal health and behavior, improving livestock management. However, these systems often lack integration with broader data ecosystems, such as weather forecasts and satellite imagery, limiting their predictive capabilities. Recent advancements in edge computing have sought to address these limitations by enabling real-time data processing at the source, reducing latency and improving efficiency [43].

The advent of autonomous farming technologies, such as self-driving tractors and drones, has further revolutionized agriculture. The authors of [44] proposed a drone-based agricultural monitoring system using Optical Camera Communication (OCC), enabling high-density sensing at low cost, with a trajectory control algorithm reducing flight time by 30 % and achieving low bit error rates in real-world experiments. Similarly, autonomous robot developed by [45] employs a CNN-based system to detect and address tomato plant diseases, autonomously activating a pesticide sprayer for precision targeting, mitigating risks associated with human error in sustainable farming practices. However, these technologies often operate in isolation, lacking integration with comprehensive farm management systems. Recent efforts have focused on developing interoperable systems that combine autonomous machinery with IoT and cloud-based platforms, as demonstrated by [46].

[2] https://www.365farmnet.com

Cloud-based decision support systems (DSS) have emerged as powerful tools for agricultural management. Platforms like *CropWise* by Syngenta provide farmers with actionable insights based on historical and real-time data, enabling optimized planting, fertilization, and harvesting [33]. Another example is *DSSAT* (Decision Support System for Agrotechnology Transfer), a tool used for simulating crop growth and yield under various environmental and management conditions. It integrates data on soil, weather, and crop management to provide insights into agricultural practices, enhancing productivity and sustainability [32]. The DSSAT model has been utilized to compare conventional and alternate wetting and drying (AWD) techniques in rice farming, demonstrating a 99.78 % accuracy in predicting grain yield and improved water use efficiency with AWD [47]. While these systems excel in data integration and visualization, they often lack extensibility, making it difficult for users to develop custom applications tailored to their specific needs. Recent advancements in predictive analytics, such as those proposed by [48], have sought to address these limitations by enabling more accurate and timely decision-making.

2.1. Comparison of technical innovations

A comparative analysis of existing smart farming platforms is presented in Table 1, highlighting key capabilities such as scalability, interoperability, extensibility, batch and real-time processing, cost-effectiveness, region-specific and governance model. *WALLeSmart* distinguishes itself through its Lambda-based architecture, modular plugin system, Walloon Agricultural DataHub, and farmer-led governance model. Unlike *FarmBeats* [23], which lacks extensibility, *WALLeSmart* supports custom DSSs and integrates platforms like Myawenet, MyCDL, and DjustConnect via a one-stop shop. Compared to *Climate FieldView* [26], a proprietary platform, *WALLeSmart* uses open-source tools (e.g., Apache Kafka, Spark) and secure protocols (e.g., JWT, CSAM login) for interoperability. The Walloon Agricultural DataHub enables secure data exchange, a feature not emphasized in *farmmaps* [27] or *DEMETER* [31]. The farmer-led governance model ensures autonomy, unlike *Barto* [28], which relies on public-private partnerships. Compared to [1], *WALLeSmart* adds real-time processing, a DataHub, and a governance model, enhancing scalability and trust.

2.2. Gaps and opportunities

While existing smart farming platforms and tools have made significant strides in optimizing agricultural practices, several gaps remain. First, many platforms are not designed to address region-specific challenges, such as the diverse climatic conditions and agricultural practices. Second, the integration of heterogeneous data sources, including IoT sensors, satellite imagery, and weather forecasts, remains a challenge due to issues of interoperability between systems and platforms. Third, while some platforms offer predictive analytics, few provide extensible frameworks for developing custom applications. Finally, user-friendly interfaces and real-time data processing capabilities are often lacking, hindering widespread adoption.

The proposed platform addresses these gaps by offering a scalable, cloud-based solution tailored to the needs of the Wallonia region of Belgium. By integrating IoT sensors, satellite imagery, and weather data, *WALLeSmart* provides real-time insights and predictive analytics to support data-driven decision-making. Its extensible plugin system allows developers and agricultural professionals to create custom applications, addressing specific farming needs such as optimized irrigation, fertilization, and pest control. Furthermore, *WALLeSmart* emphasizes user-friendly data visualization and real-time processing, making it accessible to farmers of all scales. Before detailing our proposed system, in the next section, we provide an overview of smart farming technologies and their role in transforming traditional agricultural practices.

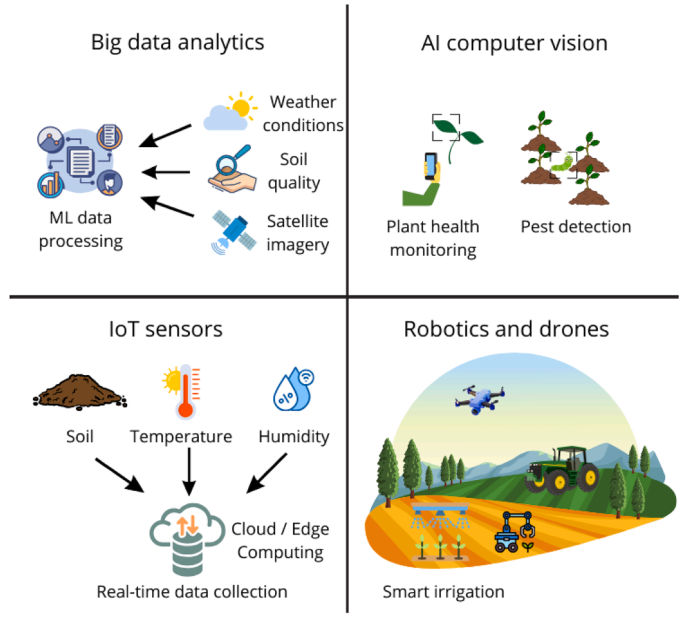


Fig. 1. Overview of smart farming technologies.

3. Smart farming technologies overview

The integration of advanced technologies into agriculture has transformed traditional farming practices, ushering in the era of smart farming. Fig. 1 presents a comprehensive overview of the core components driving this transformation. IoT sensor devices play a critical role in measuring parameters such as soil quality, temperature, and humidity, offering real-time data to monitor environmental conditions. Robotics and drones are utilized for tasks like aerial imaging, crop monitoring, and precision operations, significantly enhancing efficiency and accuracy. Real-time data collected from these diverse sources empowers farmers to make timely and informed decisions. At the heart of this ecosystem lies data processing and analytics, enabling the management and analysis of large agricultural datasets to extract actionable insights. Collectively, these technologies enable seamless data collection, processing, and analysis to advance modern agriculture.

The exponential growth of data in agriculture, driven by IoT sensors, drones, satellites, and farm machinery, necessitates robust big data processing architectures and scalable cloud computing solutions, which form the backbone of smart farming. In the next sections, we will discuss the key components of big data processing architectures and the role of cloud computing in agriculture.

3.1. Big data processing for smart farming

Big data refers to both structured and unstructured data that is too large or complex to be processed using traditional data processing tools and techniques. The term “big” is characterized by the 5 Vs: Volume, Velocity, Variety, Veracity, and Value [49]. Some companies have also introduced additional dimensions to further describe big data. To fully harness the potential of this vast amount of data, it must be effectively managed and analyzed. This process begins with ingesting and integrating raw data from diverse sources into a centralized storage repository known as a Data Lake. Next, big data solutions are employed to process the data, utilizing both long-running batch jobs and real-time processing to filter, aggregate, and prepare the data for analysis. The final step involves extracting actionable insights through data analysis and reporting.

In general, big data systems must meet two fundamental requirements: (1) the ability to handle massive real-time data streams from

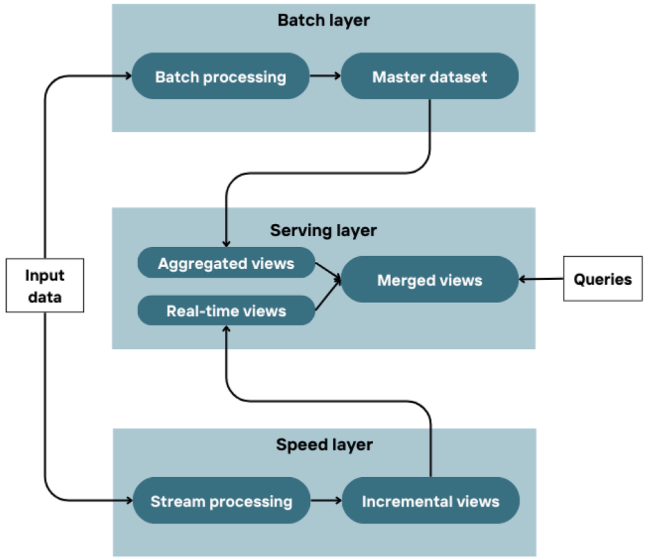


Fig. 2. *Lambda* architecture.

multiple sources, and (2) the capability to perform immediate analysis to deliver timely results [50]. To address these requirements, three state-of-the-art software architectures are commonly used: *Lambda*, *Kappa*, and a *Hybrid* that combines elements of both.

3.1.1. Lambda architecture

The Lambda architecture describes a generic, scalable, and fault-tolerant real-time data processing framework [51]. It consists of two parallel processing branches: one for batch processing and another for real-time processing. As illustrated in Fig. 2, incoming data is simultaneously fed into both branches. In the batch layer, data is appended to a centralized storage area known as the master dataset. This data is typically processed at a later stage using batch processing tools like Apache Hadoop, generating batch views. In contrast, the speed layer processes data in real time using stream processing tools such as Apache Storm, producing incremental views. Since processing large datasets in the batch layer can be time-consuming, the results are often not up-to-date. The speed layer addresses this limitation by providing access to the most recent data. Finally, the serving layer merges the outputs from both the batch and speed layers, enabling end-users to query the combined data for comprehensive insights.

3.1.2. Kappa architecture

The Kappa architecture aims to simplify the development and maintenance of data processing systems [52]. Unlike the Lambda architecture, it utilizes only one processing branch-the speed layer-as illustrated in Fig. 3. In this architecture, incoming data is processed immediately in real time using a single stream processing tool, such as Apache Kafka or Apache Flink. The results are stored as incremental views, which are then made available in the serving layer for querying and analysis. A key feature of the Kappa architecture is its ability to reprocess the entire master dataset as needed, while simultaneously handling incoming real-time data streams. This dual capability ensures flexibility and scalability. The primary advantage of the Kappa architecture lies in its single data processing engine, which simplifies system design and reduces complexity. However, this approach may require increased computational power and storage capacity to handle the parallel processing of large datasets and real-time data streams efficiently.

3.1.3. Hybrid architecture

While both the *Lambda* and the *Kappa* architectures support the processing of historical data and real-time data, they operate in separate

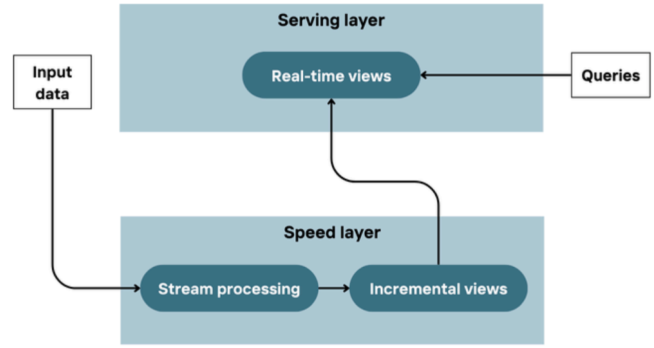


Fig. 3. *Kappa* architecture.

modes. This separation requires users to manually integrate the results from both processing modes to enable a comprehensive analysis of the data [53]. To address this limitation, [53] introduced a hybrid processing architecture called *BRAID*. *BRAID* intertwines the processing of historical and real-time data by establishing communication channels between the batch engine and the stream engine. This integration allows for automatic comprehensive analyses with minimal overhead. The authors also explored various implementation techniques for their proposed architecture. In a similar vein, the study by [54] proposed a conceptual architecture for big data streaming integrated with complex event processing (CEP), named *BiDCEP*. This system extends the *Lambda* and *Kappa* architectures to accommodate the complex event processing and event management domains of enterprise IT. The authors initiated a technical discussion on the benefits of combining these architectures and provided a motivational example to illustrate their approach.

3.2. Cloud computing for smart farming

Cloud computing is a technology that enables on-demand access to shared pools of configurable computing resources (e.g., servers, storage, networks, applications, and services) over the internet. These resources can be rapidly provisioned and released with minimal management effort or service provider interaction [55]. Cloud computing is defined by five essential characteristics, three service models, and four deployment models. The key service models include Infrastructure as a Service (IaaS), which provides virtualized computing resources; Platform as a Service (PaaS), which offers development and deployment platforms; and Software as a Service (SaaS), which delivers applications over the internet. The four deployment models are Private Cloud, Community Cloud, Public Cloud, and Hybrid Cloud, each catering to different organizational needs and use cases [55]. Many agricultural businesses adopt hybrid cloud models that combine on-premise infrastructure with cloud services. This approach offers flexibility and ensures sensitive data remains securely stored locally, while benefiting from the scalability of the cloud. Cloud computing offers numerous benefits, such as scalability, cost efficiency, and flexibility making it a valuable tool for modern agriculture. However, its adoption also introduces several challenges. Key issues include security and privacy, downtime and reliability, cost management, data transfer limitations and vendor lock-in, are the main challenges in the use of cloud, especially in the agricultural domain [56].

3.2.1. Fog computing in smart farming

While cloud computing provides centralized processing, fog computing extends cloud computing by acting as an intermediate layer between edge devices and the cloud. Fog computing is defined as a decentralized computing architecture that extends cloud computing capabilities to the edge of the network, closer to the data source. It enables data processing, storage, and analysis at the edge devices (e.g., IoT sensors, gateways) rather than sending all data to a centralized cloud. This approach reduces latency, bandwidth usage, and reliance on cloud infras-

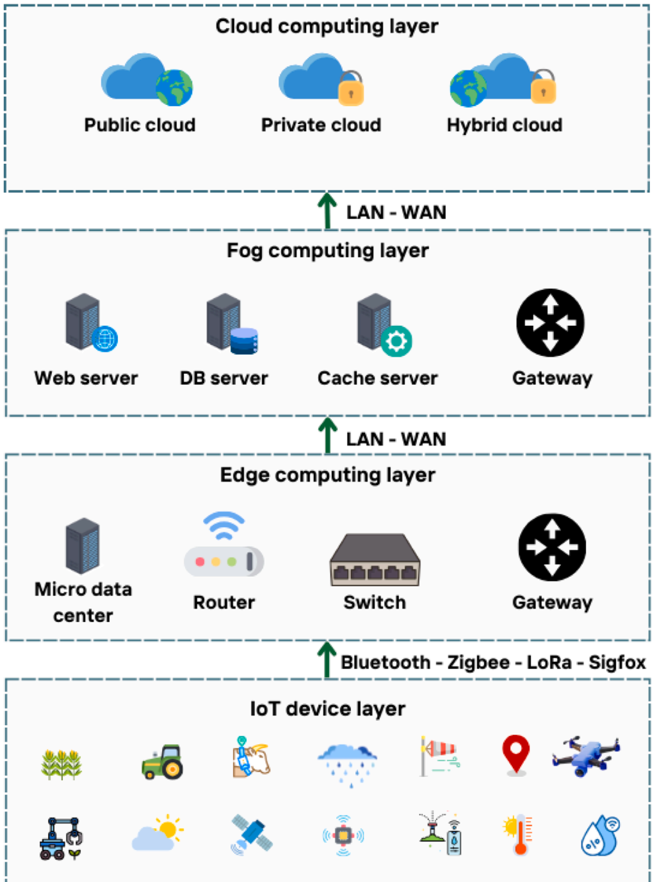


Fig. 4. Overview of cloud, fog, and edge computing layers in smart farming architecture.

tructure while supporting real-time decision-making and localized data processing [57]. Fog computing is particularly useful in applications requiring low latency, such as smart farming systems [58]. The Fog layer consists of large-scale, geo-distributed Fog nodes deployed at the edge of networks. Each Fog node is equipped with onboard computational resources, data storage capabilities, and network communication facilities, enabling it to act as a bridge between IoT devices and the cloud within the IoT ecosystem [59].

3.2.2. *Edge computing in smart farming*

Edge computing is a distributed computing paradigm that brings data processing and storage closer to the source of data generation, such as IoT devices, sensors, or local servers, rather than relying on a centralized cloud infrastructure. By processing data at the "edge" of the network, edge computing reduces latency, minimizes bandwidth usage, and enables real-time decision-making [60]. While both Fog computing and Edge computing aim to reduce latency and improve efficiency by decentralizing data processing, they differ in their architecture, scope, and applications. Fog computing uses hierarchical layer of Fog nodes (e.g., routers, gateways) enabling distributed intelligence across a network, while Edge computing emphasizes localized processing at the data source [59]. Edge computing has numerous applications in agriculture, including precise irrigation, fertilization, pest control, livestock Monitoring, and supply chain traceability [22].

3.2.3. *Serverless computing*

Serverless computing, a relatively new and rapidly evolving field in cloud computing. Serverless is a model where the cloud provider dynamically manages the allocation and provisioning of computing resources, allowing developers to focus solely on writing and deploying code without worrying about underlying infrastructure. In serverless computing, applications are broken into small, event-driven functions (e.g., AWS Lambda, Google Cloud Functions) that are executed in response to specific triggers, such as HTTP requests or database updates. The cloud provider automatically scales resources based on demand, and users are billed only for the actual execution time of the functions, making it a cost-efficient and scalable solution for modern applications [61]. Serverless architectures allow smart farming applications to run event-driven tasks without managing servers. For example, a serverless function can trigger real-time notifications when a sensor detects critical thresholds in soil moisture levels [62].

Fig. 4 depicts a multi-layered architecture for smart farming, consisting of IoT devices for data collection, edge computing for local processing, fog computing for intermediate analysis, and cloud computing for advanced analytics and storage.

3.3. *Ethical challenges in smart farming*

Smart farming raises ethical concerns, including algorithmic bias and data ownership. AI models may produce biased recommendations if trained on non-representative datasets [21]. *WALLeSmart* mitigates this through diverse training data and explainable AI. Data ownership is critical, as farmers risk losing control over sensitive data [22]. *WALLeSmart*'s consent management system, integrated with CSAM[3] login (including itsme®[4]), allows farmers to control data access, ensuring GDPR compliance and autonomy, as detailed in Section 4.7.

4. WALLeSmart system architecture

In this section, we present the architecture of the *WALLeSmart* framework. Drawing from the previous discussions, we adopt a *Lambda* architecture to implement our system. This choice is motivated by the distinct nature of batch and real-time analysis in our context, necessitating separate tools for each task. Additionally, the *Lambda* architecture ensures linear scalability, fault tolerance against hardware failures, and significant performance enhancements [51]. Our implementation comprises the three key layers of the *Lambda* architecture: the batch layer, the speed layer, and the serving layer. The tools selected for our system are based on stringent criteria, including being open-source, scalable, distributed, extensible, and widely validated in production environments. To support this architecture, we adopt a private cloud computing approach. This decision ensures greater control over data security, compliance, and resource allocation while maintaining scalability and flexibility [55]. The private cloud infrastructure is hosted on-premise, serving as the backbone for managing data storage, processing, and analytics. The following sections provide a detailed overview of the system's underlying layers and their roles in the framework.

4.1. *Master dataset*

Data flowing in the platform goes through several steps. First, data is coming from two main sources (i) dairy farms and (ii) weather station observations. Incoming data can be in various formats, such as XML, JSON, or XLS files received in different ways, such as Web service, Email attachment, or MQTT server, which all depend on the sensor manufacturer being used. Next, data is transmitted to our cloud platform via different network protocols, such as Wi-Fi, 3G/4G, or LoRa. We use EMQ

[3] CSAM (Central Authentication Service of the Belgian Government): A secure digital gateway that enables users to access various Belgian government services online. It ensures reliable identity verification and access control.

[4] itsme®: A widely used Belgian mobile identity app that allows users to authenticate themselves securely online, sign documents digitally, and give consent for data sharing. It is integrated with CSAM and recognized for its high level of security and user trust.

X as an MQTT message broker. Second, when the data arrives at our platform, it will be processed by a data ingestion tool that is responsible for capturing and storing real-time messages to be consumed by a stream/batch processing consumer. In our study, we use Apache Kafka as a data ingestion tool. It can scale to handle millions of messages per second. Third, and after capturing real-time messages, the platform processes them to gain knowledge about the data. Finally, storage tools are used as an output destination to capture real-time and batch data for archiving, or for further processing. In our architecture, we use Apache Cassandra, PostgreSQL, and MinIO for NoSQL, relational data, and unstructured data, respectively. Additionally, PostGIS is used for storing geospatial data.

4.2. Batch layer

The batch layer is responsible for managing large-scale historical data processing. We utilize Apache Spark, an open-source, distributed processing system, to perform complex computations efficiently. Spark's fault-tolerance and scalability make it ideal for generating accurate, pre-computed views of historical data that are periodically updated. This layer processes raw data ingested from diverse sources, such as IoT devices and external systems, and stores the results in a format optimized for querying.

4.3. Speed layer

The speed layer is designed to process real-time data streams with minimal latency. For this, we use Apache Kafka Streams, a lightweight yet powerful stream-processing library. Kafka Streams processes incoming data in real-time, providing quick and approximate results that are critical for immediate decision-making. This layer ensures a seamless flow of time-sensitive data into the system, complementing the batch layer by focusing on responsiveness and agility.

4.4. Serving layer

The serving layer is a critical component of the *WALLeSmart* architecture, designed to deliver insights, reporting, and data visualization in real-time. The web application of *WALLeSmart* is structured into backend and frontend parts for modularity and scalability. The backend is implemented as a Node.js API server, leveraging Google's V8 JavaScript engine for building fast and scalable network applications. It provides the core functionalities of the platform, including data processing, user authentication, role-based access control, and API endpoints. The frontend is a Vue.js application, ensuring a reactive and user-friendly interface for interacting with the platform. Farm data is visualized through interactive maps powered by OpenStreetMap, while interactive charts are rendered using the Apache ECharts library, offering intuitive and dynamic visualizations. To support researchers and developers, *WALLeSmart* provides a GraphQL Web API for accessing farm and weather data, enabling advanced analysis and modeling. GraphQL's flexible query capabilities allow users to request precise information and deliver a complete description of the data schema. Additionally, GeoServer is integrated into the platform to enable geospatial queries, providing a robust service for accessing spatial data and performing spatial analysis. The architecture of *WALLeSmart* platform is illustrated in Fig. 5.

4.5. Development of decision support application in WALLeSmart

WALLeSmart supports Decision Support Systems (DSSs) through a modular, scalable template. DSSs, labeled "WALLeSmart Ready", are hosted in an application marketplace (Fig. 9), including apps for weather forecasting, grass growth estimation (*W@llHerbe*), and electronic field logbooks. Each DSS uses a Vue.js 3 frontend (with TypeScript, Pinia, Naive UI) and a Node.js backend, ensuring modularity. Scalability is achieved via Docker containers and GitLab CI/CD pipelines, with Kafka

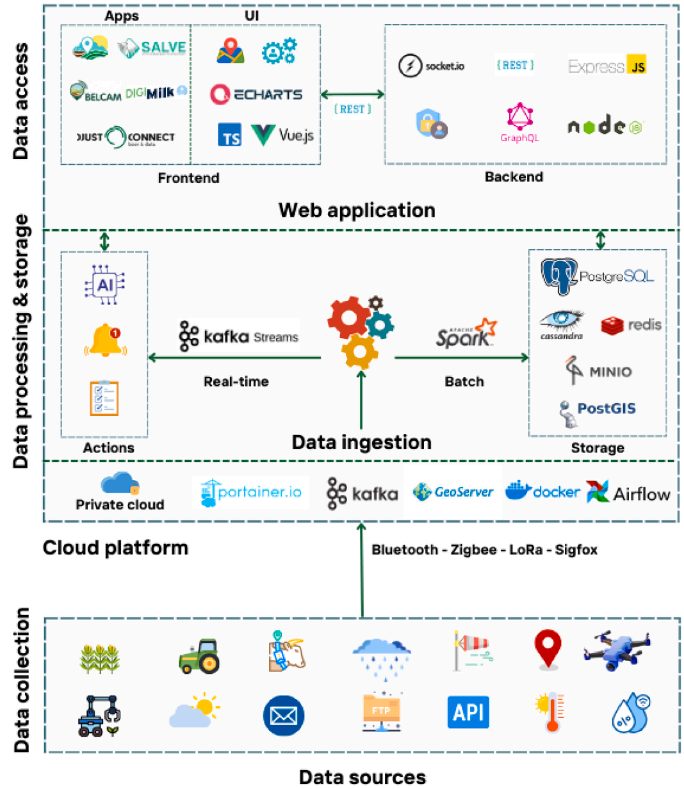


Fig. 5. *WALLeSmart* platform architecture.

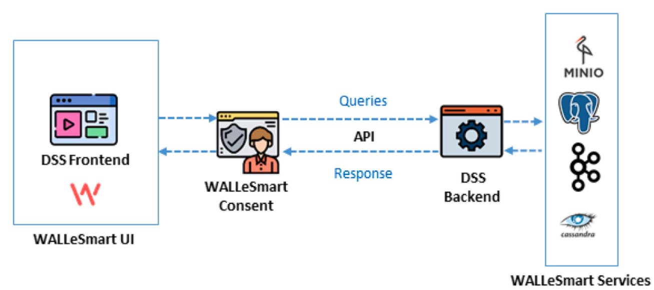


Fig. 6. Architecture of a DSS in *WALLeSmart*.

handling data streams for high workloads. The development process includes:

1. Configuring GraphQL APIs for data access (e.g., milk production, weather).
2. Integrating with MinIO, PostgreSQL, Cassandra, and PostGIS for storage.
3. Verifying user consent via CSAM login for secure data access.
4. Building and testing using modern tools (e.g., jest, vitest, pnpm/npm).
5. Deploying via Docker and automated pipelines.

Fig. 6 illustrates this pipeline, ensuring seamless integration and scalability.

4.6. Connection to external platforms

The architecture of *WALLeSmart* facilitates secure integration with public and private agricultural platforms that manage user accounts and data. At its core, the system employs secure protocols, including JSON Web Token (JWT) authentication and RSA-based cryptographic key pairs, ensuring secure and efficient communication. The connection process involves user authentication, selecting a platform, and initiating a request by providing an associated Email. *WALLeSmart* generates a JWT, sends it to the external platform, and sets the connection status to “pending". The external platform validates the JWT, verifies the Email, and updates its status. A confirmation Email is sent to the user, and upon approval, the platform notifies *WALLeSmart*, which performs final validations. After verification, the connection status updates to “connected" as shown in the workflow outlined in Fig. 7. Once connected, the platforms can securely share user data and enable advanced features such as Single Sign-On (SSO), streamlining access and enhancing functionality for users.

4.7. Security and privacy

WALLeSmart implements robust security measures:

- **Authentication:** Multi-factor authentication via CSAM login (including itsme) and JWT tokens with RSA key pairs.
- **Authorization:** Role-based access control with granular permissions.
- **Data Privacy:** End-to-end encryption (TLS 1.3) and AES-256 for data at rest.
- **Audit Logging:** Comprehensive tracking of access and changes.
- **Compliance:** GDPR adherence via consent management and the a-Box for secure document access.

All data exchanges between components are encrypted using TLS 1.3, and sensitive data is stored using AES-256 encryption at rest.

4.7.1. Ethical considerations

WALLeSmart addresses algorithmic bias through diverse training datasets (e.g., data from 30 farms of varying sizes) and explainable AI, audited regularly [22]. Data ownership is ensured via a consent management system integrated with CSAM login, allowing farmers to control data sharing with partners like ARSIA and Milk Committee (Fig. 8f). The farmer-led governance model, supported by Elevéo, UMons, CRA-W, ULiège, WalDigiFarm, ARSIA, and Milk Committee, ensures transparency and autonomy, building trust.

4.8. Generalizability across geographic regions

WALLeSmart, while tailored to the Wallonia region, is designed with broader applicability in mind. Its modular plugin architecture and the Walloon Agricultural DataHub support the integration of region-specific data sources-such as local weather APIs, crop-specific sensors, and regulatory systems-and allow customization of decision support systems (DSSs) to accommodate diverse agricultural practices, including those in Mediterranean or arid environments. The use of open-source technologies ensures portability across different cloud infrastructures. Furthermore, the platform enables cross-regional data exchange by connecting with other ecosystems, such as DjustConnect in Flanders, promoting interoperability beyond Wallonia.

5. Implementation and results

The WALLeSmart platform is designed to address the needs of the agricultural sector, with a specific focus on the dairy farming sector in the Belgium Wallonia region as a case study. The primary objectives of this case study are to help farmers (i) better manage their dairy cows, including health care, reproduction, milk production, and movement tracking, and (ii) monitor weather conditions, both historical and forecasted, in an intuitive and efficient manner. The data used in this case study is sourced from two main providers: Elevéo[5] for cow behavior data and Agromet[6] for weather data [63].

[5] https://www.awenet.be/awe/commun/asbl/asbl.php
[6] https://agromet.be/

Table 2
Agromet weather data parameters.

Parameter	Description	Unit
tsa	Air Temperature	°C
hra	Relative Humidity	%
tha	Wet Bulb Air Temperature	°C
tsf	Leaf Temperature	°C
tss	Soil Temperature at 20 cm	°C
vvt	Wind Speed at 2 m	m/s
ens	Global Solar Radiation	W/m^2
dvt	Wind Direction	°
plu	Precipitation	mm
etp	Reference Evapotranspiration	mm/day
hct	Leaf Wetting Duration	min wet

5.1. Data sources

Our system integrates data from two primary sources:

1. **Elevéo:** Collected via SenseHub[7] and SmartVel[8] sensors. The dataset spans from January 2007, to February 2025, and includes:
   - 30 dairy farms,
   - 8637 unique cows,
   - 61,130 measurement-sets, each consisting of 23 parameters (see Table 3 for details),
   - data comprises 1.5 GB.
2. **Agromet:** Collected from a network of 32 agrometeorological stations covering the entire Wallonia region. The dataset includes:
   - Historical measurements from January 1, 2008, to January 8, 2025.
   - Forecast data for the next 7 days.
   - 11 key parameters (see Table 2 for details),
   - Over 3 million historical measurements recorded hourly.
   - 6879 hourly forecast measurements,
   - totaling 3.5 GB of data.

5.2. Deployment

To ensure data sovereignty, regulatory compliance, and trust among end users, *WALLeSmart* is hosted on a private cloud infrastructure managed by the Elevéo. This setup avoids reliance on commercial public cloud providers such as AWS or Azure, offering a regionally governed and transparent alternative for data hosting. The infrastructure consists of Linux-based multi-core servers deployed in a secure data center in Wallonia, Belgium. The platform uses container-based virtualization (Docker) and follows a microservices architecture, allowing modular deployment and easy maintenance. Services are orchestrated via Docker Compose, with an ongoing transition to Kubernetes for improved elasticity and automated scaling. Importantly, farmers and agricultural stakeholders are not expected to install or maintain any physical infrastructure themselves. The platform is accessed through a responsive web interface or mobile application, requiring only a standard internet connection. IoT devices deployed on farms—such as SenseHub for livestock monitoring or SmartVel for reproductive event prediction—communicate with the platform through secure protocols (e.g., MQTT over TLS, HTTPS), enabling real-time data transmission without local storage overhead. This private infrastructure is designed with GDPR compliance at its core. A key feature is the integrated consent management system based on the Belgian CSAM authentication service and *itsme* identity verification. This allows farmers to retain full control over their data by explicitly granting or revoking access to authorized third

[7] http://www.scrdairy.com/cow-intelligence/sensehub.html
[8] https://www.evolution-xy.fr/fr/monitoring/smartvel

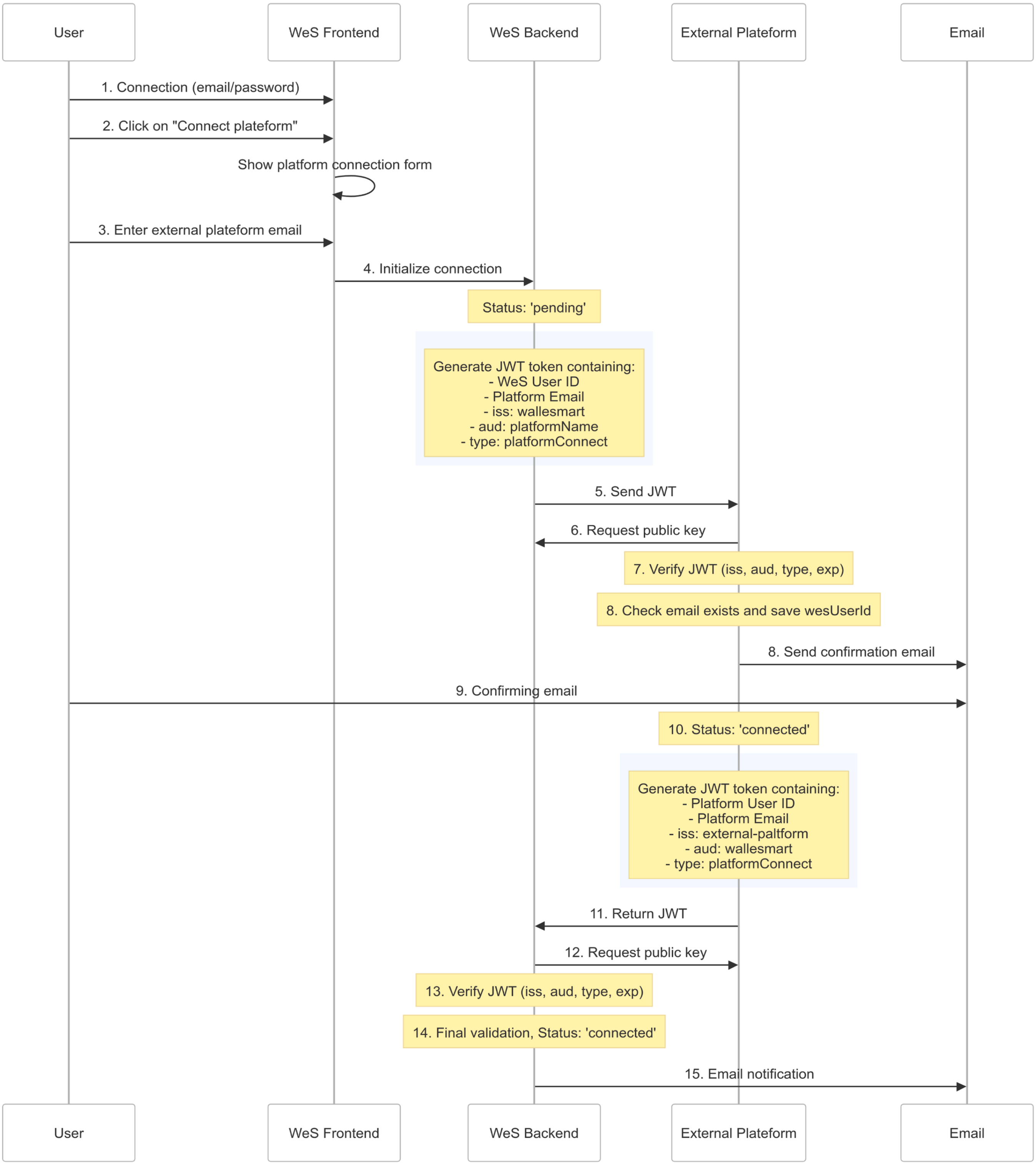


Fig. 7. Detailed workflow for secure integration of *WALLeSmart*'s user account with external platforms.

parties such as ARSIA, the Milk Committee, or research institutions. Access control policies are enforced via centralized authorization services and audited periodically to ensure transparency and accountability. Despite being hosted in a private environment, *WALLeSmart* is scalable. The system can handle increased workloads by horizontally scaling compute and storage nodes. This demonstrates the ability of the architecture to support both regional deployments and potential national or cross-border extensions.

Cost-effectiveness of the private cloud.. The "cost-effective" label in Table 1 reflects total cost-of-ownership (TCO) savings achieved by (i) reusing existing regional data center facilities and commodity hardware, (ii) avoiding public-cloud egress charges for continuous data sharing with partners (e.g., ARSIA, MilkBE), (iii) leveraging open-source components (Kafka, Spark, Cassandra, PostgreSQL, PostGIS, MinIO) with no per-seat licensing, and (iv) right-sizing capacity to predictable regional workloads. For our workload profile (continuous ingestion from 32

Table 3
Elevéo milk control data parameters.

Parameter variable	Description
BSAUMON	Official animal number
HEURE_TRAIT1_DEB	Controlling date
LAIT_24_VAL, POURC_PROT_24_VAL, POURC_MG_24_VAL	Controlled daily production of milk, protein and fat concentration
CELL_24_VAL, UREE_24_VAL	Controlled cell and urea content
JEL, JET	Lactation days, drying up days
NOLACT	Lactation number
DATE_VEL, DATE_TAR	Date of calving and date of drying up
LAIT, MG, PROT	Cumulative production of milk, protein and fat concentration
LAIT365, MG365, PROT365	Cumulative production over 365 days of milk, protein and fat
PIC	Peak production value

stations and periodic dairy batches), the private setup reduces recurring costs compared to equivalent managed services while preserving performance and data sovereignty.

5.3. The data value chain in WALLeSmart

The WALLeSmart platform follows a well-defined data value chain to ensure efficient data flow and processing. This section describes the key steps in the data value chain, including data acquisition, ingestion, processing, and visualization.

5.3.1. Data collection

The data acquisition step represents the foundational layer of the platform, responsible for collecting data from various sources. Data is collected from Elevéo and Agromet using SOAP and RESTful APIs on an hourly and daily basis. The platform supports multiple data source formats, including EMQ as the MQTT server for real-time data streams, FTP server for file-based data transfer, and Email attachments for manual data uploads. The received data is in XML and JSON formats. Data collection algorithms are implemented in Python scripts, which are scheduled for execution using Apache Airflow for task orchestration and scheduling.

5.3.2. Data ingestion

In the data ingestion step, the collected data is ingested into the platform for further processing. Python scripts from the data acquisition step act as producers, publishing data to Apache Kafka *topics*. Kafka topics are *partitions* and replicated across multiple servers (*broker*) to ensure scalability and fault tolerance. For MQTT sources, data is published directly from the EMQ server to Kafka topics.

5.3.3. Data processing

The data processing step involves real-time and batch processing of the ingested data. Data published to Kafka topics is processed using Kafka Streams for real-time stream processing. Kafka Streams applications perform transformations (e.g., date formatting) and collect statistics in real time. For batch processing, Apache Spark is used to handle tasks such as calculating daily cow movement or average temperature. Processed data is republished to Kafka topics and stored in PostgreSQL (for relational data), PostGIS (for geospatial) or Apache Cassandra (for NoSQL data) using Kafka connectors. Task scheduling and workflow management for batch processing are handled by Apache Airflow.

5.3.4. Data visualization

The final step in the data value chain is data visualization, which provides actionable insights to users through a web application. The application supports multiple user roles:

- Farmers can access Agromet weather forecasts and view a map of the Wallonia region with weather stations and current conditions (see Fig. 8b).
- Farmers can efficiently manage their data and application permissions through an intuitive, user-friendly interface. These permissions are associated with various data sources, including Elevéo and ARSIA. The interface features a comprehensive table that clearly displays the status of permissions-such as validated, expired, or pending-across different data categories. Permissions can be managed manually, by the platform, or through a connected web service, offering flexibility and control over data access.
- Administrators can manage stream and batch job scheduling using Apache Airflow (see 8c), edit user profiles, and configure system settings.
- Researchers can query the database using GraphQL API to ask complex questions, such as "*How much milk is produced by lactating cows daily?*" (see Fig. 8d).

The web application uses Apache ECharts for interactive charts and integrates with GeoServer for geospatial data visualization. Geospatial data is stored and managed using PostGIS, enabling advanced mapping and spatial analysis.

5.4. The WALLeSmart features

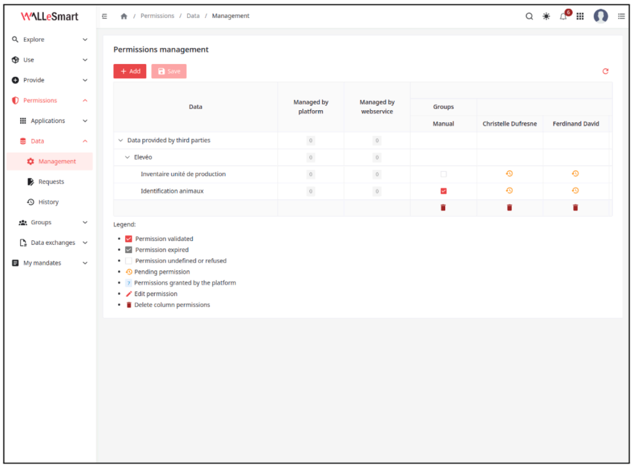
The *WALLeSmart* platform offers a wide range of features designed to meet the needs of its diverse user base. These features include:

5.4.1. Profiles and mandates managements

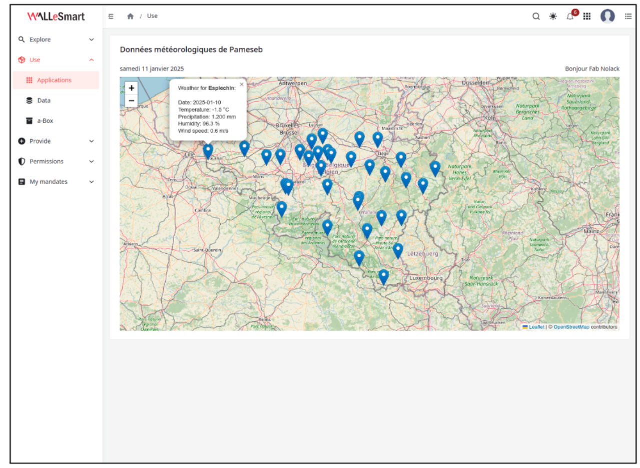
WALLeSmart provides robust user profile and mandate management capabilities. Users can create and manage profiles with specific roles and permissions, ensuring that access to data and features is tailored to individual responsibilities. This feature supports multiple user roles, such as farmers, administrators, and researchers, each with customized access levels and functionalities. Central to the system's functionality is a mandate mechanism that enables delegated access through carefully defined parameters, including principal and representative designations, feature-specific permissions, and temporal constraints. The profile and mandats management interface is intuitive and user-friendly, as shown in Fig. 8f. Additionally, the platform can automatically import user profiles using their VAT numbers, streamlining the onboarding process and reducing manual effort.

5.4.2. WALLeSmart application marketplace

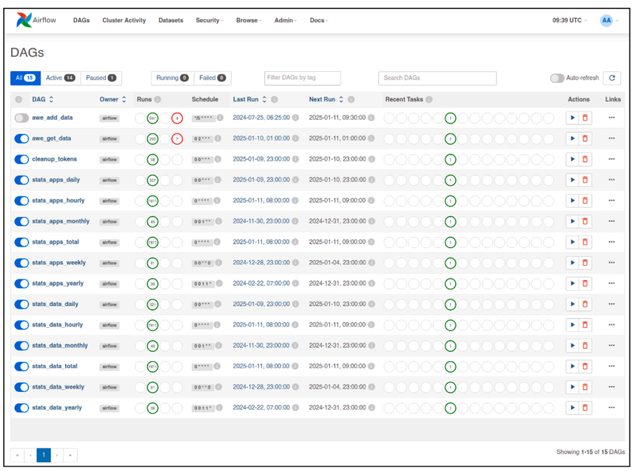
The *WALLeSmart* platform features a well-organized application marketplace that showcases various DSSs (see Fig. 9). The interface displays a clean layout with several key components, such as a navigation menu with filtering options including categories, providers, rating, and price range. Before accessing an application, users must submit a request for access. This request is reviewed by the application's developer, who can either approve or reject it based on specific criteria, such as compliance with licensing agreements. Once permission is granted, users can access the application through the "Use" menu in the *WALLeSmart* platform. This workflow ensures secure, responsible usage of applications while maintaining a user-friendly experience.



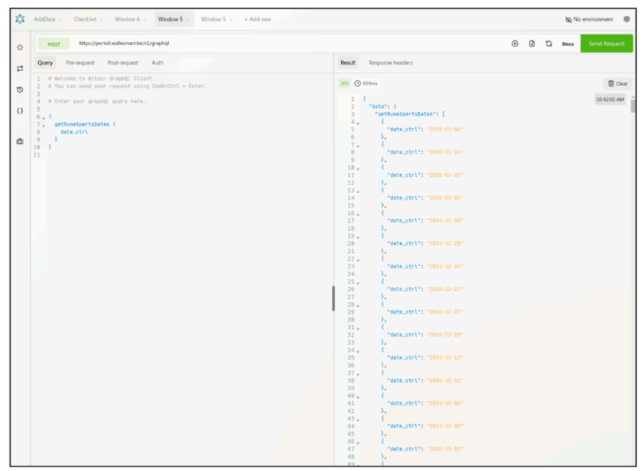
(a) Data permissions management interface in WALLeSmart.



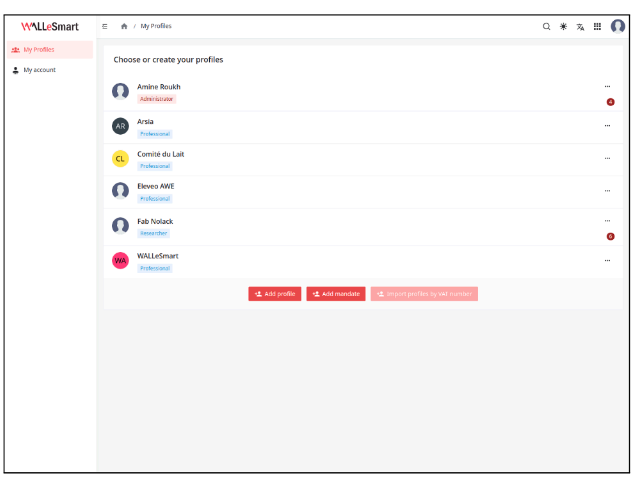
(b) Current weather conditions in Agromet application.



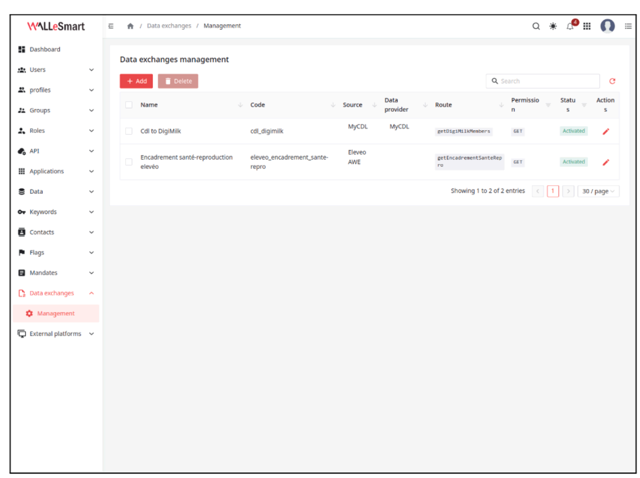
(c) Stream and batch jobs scheduling management in AirFlow.



(d) GraphQL API querying.



(e) Profiles and mandates managements.



(f) Data exchange (API) consent management.

Fig. 8. *WALLeSmart* platform and some of its features. (a) Data permissions management interface in WALLeSmart. (b) Current weather conditions in Agromet application. (c) Stream and batch jobs scheduling management in AirFlow. (d) GraphQL API querying. (e) Profiles and mandates managements. (f) Data exchange (API) consent management.

5.4.3. Data exchange consent

Fig. 8f illustrates the *WALLeSmart* data exchange management interface, designed to facilitate secure and controlled data sharing between the platform and third-party systems. The interface enables administrators to manage data exchange configurations. For instance, APIs such as “getMilkControl", belonging to third-party partners like RumeXperts, can be integrated to enable seamless data transfer. Users have the ability to approve or deny third-party access to their data via the *WALLeSmart* platform, ensuring transparency and user consent in all data-sharing activities.

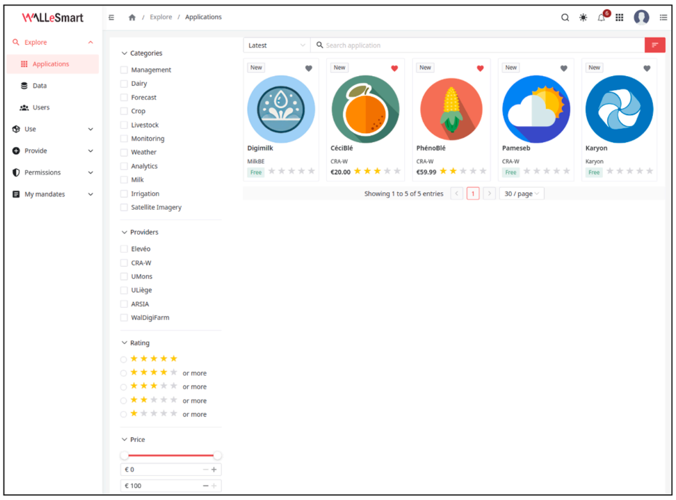


Fig. 9. *WALLeSmart* application (DSS) marketplace.

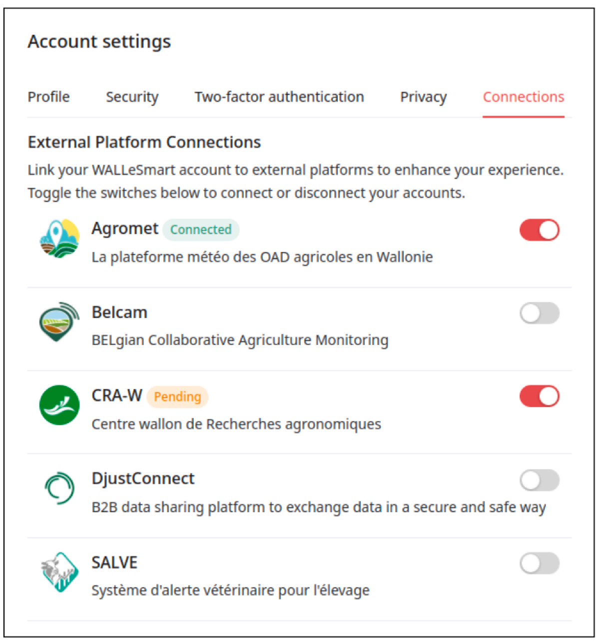


Fig. 10. External platforms management in *WALLeSmart*.

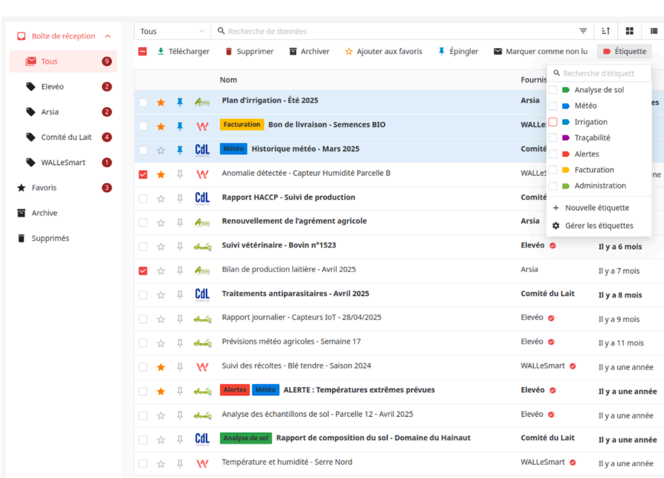


Fig. 11. a-Box interface in *WALLeSmart*, allowing centralized document access with tagging, partner filters, and notifications.

5.4.4. External platforms connection

Fig. 10 illustrates the interface for managing external platform connections within the *WALLeSmart* account settings. This feature allows users to link their account to a range of public and private platforms that manage farmer data. Platforms such as Agromet, Belcam, DjustConnect, and SALVE can be connected to enable secure data sharing and advanced integration.

Those results are shown in the speed layer through a user-friendly Web interface. This can help farmers having a closer look at weather data and timely decision making, for example, to take an estimation of irrigation needs and adjust irrigation schedules accordingly.

5.4.5. a-Box

The “a-Box” is a secure and centralized electronic mailbox integrated into *WALLeSmart*. It allows farmers to access all their official documents in one place, regardless of the sending organization (e.g., Elevéo, Comité du Lait, ARSIA). Documents are delivered via a secure GraphQL API, ensuring confidentiality and traceability. Farmers are notified of new documents through Email alerts and can manage them through an intuitive interface that supports tagging, favorites, and search functionalities. The system also provides filtering by sender or category (e.g., Météo, Facturation, Alertes), greatly simplifying information management (see Fig. 11).

5.5. Adoption and user feedback

Since its launch in 2019, *WALLeSmart* has been in a pilot phase, with 30 dairy farms (323 users, including farmers and administrators) actively using the platform as of December 2025. The platform has processed over 327k GraphQL API queries and supports 12 “WALLeSmart Ready” applications, including *SALVE*, *Agromet*, and *MyFieldBook*. User interface feedback reported here is testimonial (pilot feedback, workshops, informal interviews). Adoption barriers include:

- **Connectivity:** Slow rural internet, mitigated by offline caching and planned edge nodes.
- **Privacy:** Concerns addressed for data consent management.
- **Training:** Limited expertise, countered by workshops and intuitive interfaces.

Partnerships with WalDigiFarm, ARSIA, and Milk Committee, supported by the Walloon Region and Digital Agency, drive adoption through outreach and training.

5.6. Performance evaluation

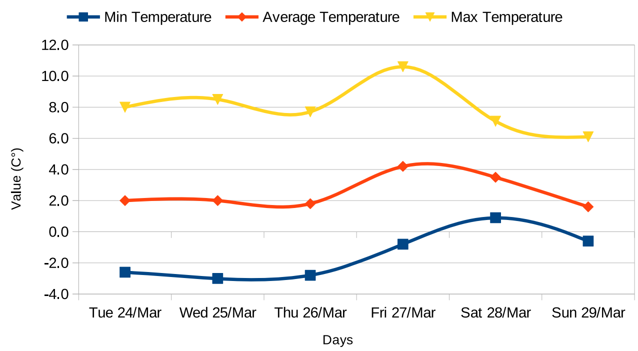
This section presents the experiments and results conducted to evaluate the effectiveness of our platform. We focus on both real-time and batch processing tasks, highlighting their performance and outcomes.

5.6.1. Real-time job execution

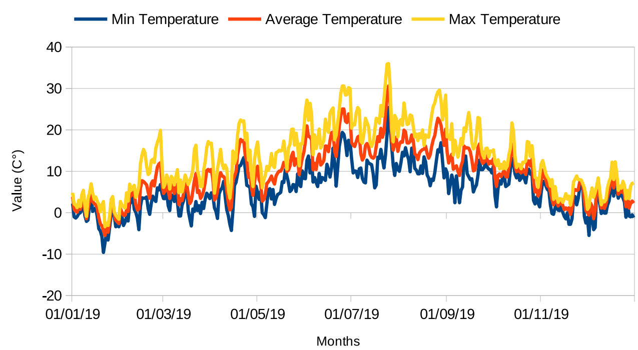
Fig. 12a illustrates the results of a real-time job executed by the speed layer. The task involves calculating the average, maximum, and minimum daily temperatures for the next seven days. Since the platform receives estimated temperature data on an hourly basis, the real-time layer aggregates this data to produce daily forecasts. The data stream is ingested into Kafka topics and processed using Kafka Streams, which performs transformations and computations in real time. The results are then republished to Kafka and stored in Cassandra. This pipeline ensures efficient and timely processing of real-time data.

5.6.2. Batch job execution

Fig. 12b showcases the results of a batch job executed by the batch layer. The task involves calculating the average, maximum, and minimum daily temperatures for the entire year of 2019 across all weather stations. This batch process is scheduled to run at regular intervals and



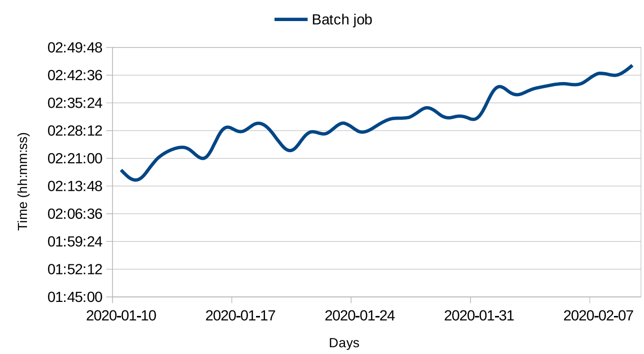
(a) Real-time job for calculating temperature forecasting for the next 7 days.



(b) Batch job for calculating temperature history for 2019.

Fig. 12. Examples of batch and real-time jobs for scheduled one weather station. (a) Real-time job for calculating temperature forecasting for the next 7 days. (b) Batch job for calculating temperature history for 2019.

(a) The execution time of a real-time job for weather data over one month.



(b) The execution time of a batch job for dairy data over one month.

Fig. 13. Performance of batch and real-time job execution. (a) The execution time of a real-time job for weather data over one month. (b) The execution time of a batch job for dairy data over one month.

operates on the entire master dataset. Leveraging Apache Spark, the batch layer accelerates computation, and the results are saved to the database upon completion.

5.6.3. Execution time analysis

To further evaluate the performance of our architecture, we measured the execution times for both real-time and batch jobs over an extended period.

- **Real-Time Job Execution Time:** Fig. 13a shows the execution time of the real-time job over a one-month period. The average execution time is consistently low, averaging 1 minute and 18 seconds per job. This efficiency is attributed to the lightweight and scalable nature of Apache Kafka Streams, which processes data streams with minimal latency. The reported time includes data ingestion, processing, and storage.
- **Batch Job Execution Time:** Fig. 13b depicts the execution time of a batch job analyzing dairy farm data. The average execution time is significantly longer, averaging 2 hours and 31 minutes. The curve exhibits an upward trend over time, reflecting the increasing volume of dairy data processed daily. To address this, the system can be scaled horizontally by adding more server nodes to distribute the computational load and improve performance.

5.6.4. Resource consumption and performance trade-offs

Resource usage was measured on a 16-core Linux server (64 GB RAM). Table 4 shows usage for real-time (weather) and batch (dairy) jobs. Real-time jobs average 2 CPU cores and 4 GB RAM, prioritizing low latency (1 minute 18 seconds). Batch jobs use up to 8 CPU cores and 16 GB RAM, optimizing throughput but increasing time (2 hours 31 minutes). Horizontal scaling mitigates bottlenecks, though it raises costs. Fig. 13 visualizes execution times and resource usage, highlighting trade-offs.

Table 4
Resource Consumption for WALLeSmart Jobs Abbreviations: RT = Real-time, CPU = CPU Cores, Mem = Memory, Disk = Storage Disk, pts = data points, s = second.

Job	Workload	CPU	Mem (GB)	Disk (GB)
RT (Weather)	1000 pts/s	2	4	0.5
RT (Weather)	10,000 pts/s	4	8	1.0
Batch (Dairy)	1000 pts	4	8	2.0
Batch (Dairy)	10,000 pts	8	16	5.0

These experiments demonstrate the platform's ability to handle both real-time and batch processing tasks effectively. While real-time jobs exhibit low and consistent execution times, batch jobs require optimization for scalability as data volumes grow. Future improvements include parallelizing computations and enhancing resource allocation to further reduce batch job execution times.

5.7. Decision sems

In this section, we present several Decision Support Systems (DSS) that leverage the diverse data ecosystem provided by the *WALLeSmart* platform. These applications illustrate the platform's ability to integrate heterogeneous data sources-ranging from weather stations and livestock records to satellite imagery and governmental databases-to support evidence-based agricultural decision-making.

5.7.1. SALVE

SALVE[9] is a web-based Veterinary Alert System that automates the analysis of livestock-related data to improve decision-making and reduce reaction times. It currently processes data from approximately 100 dairy and beef farms in Wallonia, including herd composition, milk quality, and climatic conditions. Through integration with *WALLeSmart*'s real-time data streams-especially milk quality and production metrics-*SALVE* leverages time-series analyses, multivariate covariation (e.g., somatic cell count versus urea or fatty acids), and population-level metrics (infection and cure rates) to assess animal health, detect early warning signs, and quantify environmental stressors. Alerts are generated 30 to 45 days in advance of critical issues, allowing farmers to act preemptively. This results in enhanced animal welfare, increased livestock longevity, and reduced production costs.

5.7.2. W@llHerbe

W@llHerbe is a pasture monitoring application that exemplifies the platform's Earth observation capabilities. It combines field-level grass growth measurements, Sentinel-2 satellite imagery, and biophysical growth models to estimate biomass and forage quality in near real-time. These data streams are processed through *WALLeSmart*'s infrastructure to generate localized, actionable insights for farmers.

Predictive models used. The W@llHerbe system relies on a multi-source predictive framework that combines the *ModVege* process-based grass growth model [64] with observations from the Copernicus Sentinel-1 and Sentinel-2 missions. ModVege simulates daily grass biomass dynamics based on meteorological inputs, soil characteristics, vegetation type, and management practices such as cutting, grazing, and fertilisation. To enhance prediction accuracy, satellite-derived biophysical variables-particularly the Leaf Area Index (LAI) obtained from Sentinel-2-are assimilated into the model to correct simulated biomass trajectories. Sentinel-1 data further support the detection of mowing events. Field measurements (biomass, canopy height, floristic composition) are used to calibrate and validate both the growth model and the remote-sensing estimations. This coupled modelling-EO approach provides spatially explicit, daily biomass estimates at the parcel scale, achieving significantly improved performance (RMSE $\approx$ 500 kg DM/ha) compared to using ModVege alone.

The tool is paired with a digitized grazing calendar and supports plot-level decisions such as optimized mowing schedules and feed rationing. By exploiting spatial and temporal data fusion, *W@llHerbe* enhances the sustainable use of forage resources while addressing agronomic challenges posed by increased climate variability. It will be distributed through the *WALLeSmart* app store to facilitate user adoption and ensure seamless integration with other services on the platform.

5.7.3. Digimilk

DigiMilk[10] is a digital service developed by MilkBE and hosted on the *WALLeSmart* platform as part of the updated Sustainability Monitor for Belgium's dairy sector. It enables farmers to easily access, manage, and update sustainability-related information across more than 70 initiatives grouped under nine thematic areas (e.g., animal welfare, environmental footprint, energy use). By leveraging automatic data ingestion from authoritative sources such as the Walloon Government, Arsia, and CdL, *DigiMilk* significantly reduces administrative overhead. Farmers retain full control of their data and can grant access to stakeholders such as MilkBE or dairy processors via secure, consent-based sharing mechanisms. The system illustrates how *WALLeSmart* facilitates interoperability and data governance across a fragmented agri-food ecosystem.

These DSS implementations underscore the platform's flexibility and scalability beyond the initial use of weather and milk data. In particular, *W@llHerbe* demonstrates ongoing efforts in satellite-based crop and pasture monitoring, opening avenues for future extensions in pest detection and yield forecasting. Furthermore, we clarify that the 5 GB dataset cited earlier corresponds to a single case study. For comparison, a single Sentinel-2 acquisition can reach 5GB of optical data. With an average of 60 acquisitions per year per target region, this results in 300GB/year solely from satellite imagery-highlighting the system's capacity to process data at a scale consistent with modern big data paradigms in agriculture.

5.8. Limitations

The *WALLeSmart* platform, while pivotal for advancing sustainability in Wallonia's agricultural sector, faces several limitations that hinder its adoption and performance. These challenges can be grouped into operational constraints at the farm level and technical hurdles within the platform itself.

5.8.1. Operational constraints

Slow connectivity at farm offices, a common issue in rural areas, delays data uploads and interactions with the platform, frustrating users who rely on timely updates. Similarly, the absence of power in fields limits the deployment of IoT devices essential for real-time data collection, leaving gaps in on-site monitoring. This problem is compounded by a lack of connectivity in fields, where poor or nonexistent network coverage prevents seamless data transmission from sensors. Additionally, precision mapping suffers due to limited sensor availability, restricting the platform's ability to generate detailed, accurate sustainability metrics-such as soil health or emissions-that farmers need to meet sector-wide goals.

5.8.2. Technical limitations

Technical limitations further challenge *WALLeSmart*'s performance as it integrates new data sources and scales across the region. Combining diverse datasets, such as weather patterns and dairy farm records, requires manual calibration to align heterogeneous data formats, slowing processing times and introducing potential inconsistencies. Scalability poses another concern, as batch processing performance declines when the number of connected IoT nodes increases, reducing efficiency as adoption grows. Additionally, the platform's plugin system, though versatile, presents user adoption barriers due to its complexity, demanding a level of technical proficiency that many farmers lack and requiring substantial training to ensure effective utilization.

6. Conclusions and future work

Agriculture remains a vital sector, directly impacting global food security and environmental sustainability. As agricultural challenges intensify due to climate variability, resource constraints, and increasing demand, the ability to manage complex agricultural ecosystems efficiently has never been more critical. The richness of big data generated by emerging agri-technologies offers unprecedented opportunities to address these challenges, ensuring a secure and healthy future for both people and the planet.

In this paper, we introduced *WALLeSmart*, a novel cloud-based smart farming management platform designed to harness the power of big data and cloud computing in agriculture. We began by investigating state-of-the-art platforms, big data and cloud architecture initiatives, which informed the design of our solution. Drawing inspiration from the *Lambda* architecture, we proposed a robust framework for the acquisition, processing, storage, and visualization of real-time agricultural data. Our platform integrates advanced technologies for real-time stream processing, such as Apache Kafka Streams, and batch processing using Apache Spark, ensuring scalability and efficiency.

An initial prototype of the platform was developed and tested with 30 dairy farms and 32 weather stations in the Wallonia region of Belgium. The backend demonstrated robust data management capabilities,

[9] https://www.salve.vet/
[10] https://www.digimilk.be/

while the frontend received positive testimonial feedback for its user-friendly interface and intuitive design; a formal usability assessment remains future work. We also presented examples of batch and real-time job execution, showcasing the platform's ability to handle diverse data processing tasks effectively.

While this study presents a scalable smart farming solution, several promising directions for future research remain. We plan to develop and integrate more decision-making applications into the platform, further enhancing its ability to provide actionable insights to farmers. The development of mobile applications for on-the-go access is also in the pipeline. Integrating advanced AI/ML techniques will further enhance predictive analytics and optimize resource management. Similarly, combining edge computing with cloud systems could enable real-time decision-making, crucial for time-sensitive operations. Further exploration of geospatial analytics, incorporating satellite and drone imagery, may yield more precise agricultural insights. Additionally, establishing interoperability frameworks and standardizing data exchange protocols can foster greater collaboration across platforms. Finally, advancing climate-resilient farming systems is critical to tackling the challenges posed by climate change and ensuring agricultural sustainability. These avenues highlight the immense potential for further innovation in smart farming systems to support global food security and environmental stewardship.

Declaration of generative AI and AI-assisted technologies in the writing process

In the preparation of this work, the authors utilized ChatGPT 3.5 (OpenAI, USA) to assist with grammar refinement. Following its use, the authors thoroughly reviewed and revised the content to ensure accuracy, clarity, and alignment with the intended message. The authors take full responsibility for the final content and conclusions presented in this publication.

CRediT authorship contribution statement

Amine Roukh: Writing – original draft; **Saïd Mahmoudi:** Writing – review & editing.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the Wallonia region of Belgium (SPW) under grant DGO3-F114PP011, project 2019–2025.

References

[1] A. Roukh, F.N. Fote, S.A. Mahmoudi, S. Mahmoudi, Big data processing architecture for smart farming, Procedia Comput. Sci. 177 (2020) 78–85.
[2] United Nations Department of Economic and Social Affairs, Population Division, World Population Prospects 2024: Summary of Results, 2024, https://population.un.org/wpp/Publications/.
[3] Food, A.O. o. t.U. Nations, The State of the World's Land and Water Resources for Food and Agriculture: Systems at Breaking Point, FAO, Rome, Rome, 2022.
[4] Intergovernmental Panel on Climate Change, Climate Change 2023: Synthesis Report, 2023, https://www.ipcc.ch/report/sixth-assessment-report-synthesis-report/.
[5] L. Zhang, et al., Precision agriculture: an overview, Agric. Syst. 196 (2022) 103261.
[6] S. Wolfert, L. Ge, C. Verdouw, M.-J. Bogaardt, Big data in smart farming–a review, Agric. Syst. 153 (2017) 69–80.
[7] X. Li, et al., Smart farming technologies: a review, Sensors 22 (6) (2022) 2044.
[8] H. Wang, et al., Sustainable practices in smart farming, J. Clean. Prod. 402 (2023) 136340.
[9] X. Gao, et al., Real-time monitoring and predictive analytics in agriculture, Precis. Agric. 24 (2023) 1144–1160.
[10] V. Kumar, K.V. Sharma, N. Kedam, A. Patel, T.R. Kate, U. Rathnayake, A comprehensive review on smart and sustainable agriculture using IoT technologies, Smart Agric. Technol. (2024) 100487.
[11] B. Bhushan, et al., Smart livestock monitoring: IoT in the ranching sector, Smart Agric. J. 8 (2022) 112–125.
[12] K. Jha, et al., Advances in IoT-based agriculture: an overview, Int. J. Agric. Technol. 29 (4) (2021) 1753–1770.
[13] A.T. Journal, The future of autonomous tractors in agriculture, Autonomous Technol. Agric. (2023). https://autonomous-agriculture.com.
[14] J. Wan Akashah Wan, et al., Spray performance case study of an agriculture drone in pineapple plantation, Int. J. Eng. Trends Technol. 72 (5) (2024) 275–181.
[15] A. Khanna, S. Kaur, Evolution of internet of things (IoT) and its significant impact in the field of precision agriculture, Comput. Electron. Agric. 157 (2019) 218–231.
[16] S. Sonka, et al., The role of cloud computing in data-Driven agriculture, Agric. Syst. 200 (2023) 103440.
[17] N. Perter, The internet of things report, Bus. Insider Intell. (2020). https://store.businessinsider.com/products/the-internet-of-things-report.
[18] X. Liu, N. Iftikhar, X. Xie, Survey of real-time processing systems for big data, in: Proceedings of the 18Th International Database Engineering & Applications Symposium, 2014, pp. 356–361.
[19] T. Bayer, et al., AI And machine learning in smart farming: predictive models for crop growth, Precis. Agric. J. 31 (2023) 367–382.
[20] D. Crichton, et al., Cloud computing for scalable agriculture data processing, Comput. Electron. Agric. 194 (2022) 106338.
[21] N.C. Eli-Chukwu, Applications of artificial intelligence in agriculture: a review, Eng., Technol. Appl. Sci. Res. 9 (4) (2019).
[22] X. Zhang, Z. Cao, W. Dong, Overview of edge computing in the agricultural internet of things: key technologies, applications, challenges, IEEE Access 8 (2020) 141748–141761.
[23] D. Vasisht, Z. Kapetanovic, J. Won, X. Jin, R. Chandra, S. Sinha, A. Kapoor, M. Sudarshan, S. Stratman, Farmbeats: an IoT platform for data-Driven agriculture, in: 14Th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17), 2017, pp. 515–529.
[24] E. Poornima, N. Kasiviswanath, C.S. Bindu, ADaaS: a secure attribute-based group signature-based agri-cloud framework, Int. J. Adv. Intell. Paradigms 25 (3–4) (2023) 264–279.
[25] A. Kaloxylos, A. Groumas, V. Sarris, L. Katsikas, P. Magdalinos, E. Antoniou, Z. Politopoulou, S. Wolfert, C. Brewster, R. Eigenmann, et al., A cloud-based farm management system: architecture and implementation, Comput. Electron. Agric. 100 (2014) 168–179.
[26] C. FieldView, Maximize Results with Our Digital Farming Solution, 2025, Accessed: 25-Feb-2025, https://climate.com.
[27] T.H. Been, C. Kempenaar, F.K. van Evert, I.E. Hoving, G.J.T. Kessel, W. Dantuma, J.A. Booij, L.P.G. Molendijk, F.D. Sijbrandij, K. van Boheemen, Akkerweb and farmmaps: development of open service platforms for precision agriculture, in: Precision Agriculture: Modelling, Springer, 2023, pp. 269–293.
[28] Barto, La solution intelligente pour les paysans et les paysannes, 2025, Accessed: 25-Feb-2025, https://www.barto.ch.
[29] Y. Zhang, L. Wang, Cropsense: machine learning for precision agriculture, Comput. Electron. Agric. 193 (2022) 106712.
[30] C. Cambra Baseca, S. Sendra, J. Lloret, J. Tomas, A smart decision system for digital farming, Agronomy 9 (5) (2019) 216.
[31] I. Roussaki, K. Doolin, A. Skarmeta, G. Routis, J.A. Lopez-Morales, E. Claffey, M. Mora, J.A. Martinez, Building an interoperable space for smart agriculture, Digit. Commun. Networks 9 (1) (2023) 183–193.
[32] G. Hoogenboom, C.H. Porter, K.J. Boote, V. Shelia, P.W. Wilkens, U. Singh, J.W. White, S. Asseng, J.I. Lizaso, L.P. Moreno, et al., The DSSAT crop modeling ecosystem, in: Advances in Crop Modelling for a Sustainable Agriculture, Burleigh Dodds Science Publishing, 2019, pp. 173–216.
[33] Cropwise, Cropwise agronomic digital sustainable solutions, 2025, Accessed: 25-Feb-2025, https://www.cropwise.com.
[34] P. Kunal, Big data in the bigger world of agriculture today, IEEE India Info (2019).
[35] A. Kamilaris, A. Kartakoullis, F.X. Prenafeta-Boldú, A review on the practice of big data analysis in agriculture, Comput. Electron. Agric. 143 (2017) 23–37.
[36] M.J. O'Grady, G.M.P. O'Hare, Modelling the smart farm, Inf. Process. Agric. 4 (3) (2017) 179–187.
[37] N. Tantalaki, S. Souravlas, M. Roumeliotis, Data-Driven decision making in precision agriculture: the rise of big data in agricultural systems, J. Agric. Food Inf. 20 (4) (2019) 344–380.
[38] R. Finger, S.M. Swinton, N. El Benni, A. Walter, Precision farming at the nexus of agricultural production and the environment, Annu. Rev. Resour. Economics 11 (1) (2019) 313–335.
[39] A. Weersink, E. Fraser, D. Pannell, E. Duncan, S. Rotz, Opportunities and challenges for big data in agricultural and environmental analysis, Annu. Rev. Resour. Econ. 10 (2018) 19–37.
[40] M. Wu, J. Xiong, R. Li, A. Dong, C. Lv, D. Sun, A.E. Abdelghany, Q. Zhang, Y. Wang, K.H.M. Siddique, et al., Precision forecasting of fertilizer components' concentrations in mixed variable-rate fertigation through machine learning, Agric. Water Manage. 298 (2024) 108859.
[41] T.A. Wild, L. van Schalkwyk, P. Viljoen, G. Heine, N. Richter, B. Vorneweg, J.C. Koblitz, D.K.N. Dechmann, W. Rogers, J. Partecke, et al., A multi-species evaluation of digital wildlife monitoring using the sigfox IoT network, Anim. Biotelem. 11 (1)

(2023) 13.
[42] Y. Feng, H. Niu, F. Wang, S.J. Ivey, J.J. Wu, H. Qi, R.A. Almeida, S. Eda, Q. Cao, Socialcattle: IoT-based mastitis detection and control through social cattle behavior sensing in smart farms, IEEE Internet Things J. 9 (12) (2021) 10130–10138.
[43] R. Kumar, P. Singh, Edge computing for real-Time agricultural data processing, IEEE Internet Things J. 10 (2) (2023) 1456–1468.
[44] S. Kondo, N. Yoshimoto, Y. Nakayama, Farm monitoring system with drones and optical camera communication, Sensors 24 (18) (2024) 6146.
[45] H.M. Rai, M. Chauhan, H. Sharma, N. Bhardwaj, L. Kumar, Agribot: smart autonomous agriculture robot for multipurpose farming application using IOT, in: Emerging Technologies for Computing, Communication and Smart Cities: Proceedings of ETCCS 2021, Springer, 2022, pp. 491–503.
[46] T. Brown, M. Green, Interoperable systems for autonomous farming, Agricultural Robotics 8 (3) (2022) 78–92.
[47] S. Ashwini, N. Sakthivel, S. Pazhanivelan, K. Ramah, P. Janaki, V. Ravichandran, N.S. Sudarmanian, Evaluating rice yield and resource efficiency: DSSAT analysis of conventional vs. AWD techniques in coimbatore, Plant Sci. Today 11 (4) (2024) 450–462.
[48] J. Smith, K. Johnson, Predictive analytics for smart farming, J. Agric. Informatics 14 (1) (2023) 45–60.
[49] A. De Mauro, M. Greco, M. Grimaldi, A formal definition of big data based on its essential features, Lib. Rev. (2016).
[50] M. Feick, N. Kleer, M. Kohn, Fundamentals of real-Time data processing architectures lambda and kappa, SKILL 2018-Studierendenkonferenz Informatik (2018).
[51] N. Marz, How to beat the CAP theorem, Online article, July (2011). http://nathanmarz.com/blog/how-to-beat-the-cap-theorem.HTML.
[52] J. Kreps, Questioning the lambda architecture, Online article, July (2014) 205.
[53] C. Giebler, C. Stach, H. Schwarz, B. Mitschang, BRAID, in: Proceedings of the 7Th International Conference on Data Science, Technology and Applications, SCITEPRESS-Science and Technology Publications, Lda, 2018, pp. 294–301.
[54] E. Hadar, BIDCEP: A vision of big data complex event processing for near real time data streaming, in: CEUR Workshop Proceedings, 1600, 2016.
[55] P. Mell, T. Grance, The NIST definition of cloud computing, National Institute of Standards and Technology (NIST) Special Publication 800-145 (2011).
[56] E. Symeonaki, K.G. Arvanitis, D.D. Piromalis, Review on the trends and challenges of cloud computing technology in climate-Smart agriculture, HAICTA (2017) 66–78.
[57] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, 2012, pp. 13–16.
[58] T.-C. Hsu, H. Yang, Y.-C. Chung, C.-H. Hsu, A creative IoT agriculture platform for cloud fog computing, Sustain. Comput.: Informatics Syst. 28 (2020) 100285.
[59] Y. Kalyani, R. Collier, A systematic survey on the role of cloud, fog, and edge computing combination in smart agriculture, Sensors 21 (17) (2021) 5922.
[60] W. Shi, J. Cao, Q. Zhang, Y. Li, L. Xu, Edge computing: vision and challenges, IEEE Internet Things J. 3 (5) (2016) 637–646.
[61] G. Adzic, R. Chatley, Serverless computing: economic and architectural impact, in: Proceedings of the 2017 11Th Joint Meeting on Foundations of Software Engineering, 2017, pp. 884–889.
[62] E.J.R. Campos, Edgeless: Design and Implementation of Serverless Computing at the Edge for Performance in Precision Agriculture, Master's thesis, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, 2023. Available at https://www.msjansen.com/assets/pdf/education/2023-ecampos-msc_thesis.pdf.
[63] D. Rosillon, J.P. Huart, T. Goossens, M. Journée, V. Planchon, The agromet project: a virtual weather station network for agricultural decision support systems in wallonia, south of belgium, in: Ad-Hoc, Mobile, and Wireless Networks: 18th International Conference on Ad-Hoc Networks and Wireless, ADHOC-NOW 2019, Luxembourg, Luxembourg, October 1–3, 2019, Proceedings 18, Springer, 2019, pp. 560–565.
[64] F. Piseddu, R. Martin, E. Movedi, F. Louault, R. Confalonieri, G. Bellocchi, Simulation of multi-Species plant communities in perturbed and nutrient-Limited grasslands: development of the growth model modvege, Agronomy 12 (10) (2022) 2468. https://doi.org/10.3390/agronomy12102468