Deep Science Publishing
https://doi.org/10.70593/978-81-988918-6-0_4

Chapter 4: Implementing scalable cloud computing solutions for agricultural data management

4.1. Introduction

While data is one of the key drivers of the agriculture sector, at the same time managing data from various data sources, formats and data ownerships is one of the biggest challenges for the sector. However, there are also great opportunities to explore and exploit the different sources that are being made available by public and private sector organizations to provide insights for enhancing business and economic growth in the sector as well as ensuring a secure, economic and sustainable food system. The emergence of the cloud computing model has the potential to address these challenges by enabling wide access to storage and processing capabilities and economies of scale that would be prohibitive for individual stakeholders to acquire. With the emergence of new capabilities in data management, data processing, dynamic computing and cost aware computing leadership for primary producers, data providers, and consulting services can use these capabilities to implement easy to use, fast, integrated, and scalable data management processes to support analytics needs for the agriculture sector (Jayaraman et al., 2016; Kamilaris et al., 2016; Brewster et al., 2017). Cloud Computing as Technology is an enabler for the e-Agriculture Strategy, providing a new model of data management. Cloud computing is about managing the physical infrastructure in a different way and ensuring security. The way to unlocking the potential of the cloud is by focusing on solutions that support the business process. Cloud computing represents a paradigm shift in the information technology ambiance that delivers services—computing, storage, software, and network—over the Internet, hence the name cloud. While not a new concept in IT, however, this latest generation of distributed, Internet-based services can deliver for industries around the world significant operational efficiencies, new business models, and revenue streams (Sundmaeker et al., 2016; Wolfert et al., 2017).

4.1.1. Purpose and Scope of the Study

Cloud computing has a variety of models and services that range from low-level data locality and storage services to complete virtualized environments for data analysis. One of the main advantages of cloud computing is its scalable aspects. By using cloud services, applications can respond easily to variable levels of load and suddenly become very expensive to run, because customers only need to pay for what they consume. This study presents a case for developing data services-oriented cloud tools to manage and analyze agricultural data. Agricultural output is highly variable through both time and space. Changes in human practices as well as disease and pest outbreaks can affect behavior on a global scale. Issues in exterior markets can affect production locally. In addition, there is a need to develop and promote low-cost tools that can enable small producers to connect to extensive marketing networks. As such, demand for agricultural data is expanding rapidly from local up to global levels.

Three case studies on implementing scalable services-oriented cloud computing solutions are presented to discuss the feasibility of cloud computing in agricultural data exchange, curation, and analysis: A local demand prediction tool for alley-cropped hazelnut production, a cloud-based service that manages satellite-based time-series phenological data for forested areas, and a framework that receives and curates low-cost sensor data from a global network of smart agriculture production facilities. Building on the importance of data standardization, each use case presents data standardization strategies. Analysis of the three studies brings insights not only about cloud solutions for agricultural data management, but also agricultural data management as a process. Implementing these services as web-focused uncomplicated tools with an intuitive user experience provides agriculturalists with greater freedom to explore data-sharing opportunities. This represents a step towards a future where small-holder farmers have better access to market information and where cloud computing enables the democratization of the data collection process, and where low-cost data collection networks made through crowd-sourcing can be developed and sustained.

4.2. Overview of Cloud Computing

This chapter has presented the problem of deficient management of agricultural data of increasing size that mainly occurs at national, regional, and local levels. A solution to this problem is the use of cloud computing technologies that allow an integrative horizontal solution for organizing, storing, and analyzing agricultural databases. But there are questions that need to be clarified in order to develop a suitable solution based on cloud technologies. Thus, which are the main concepts and terms whose meaning must be known to talk about cloud computing? What are the categories of cloud services? If the category of services to be used is clarified, how will these services benefit agricultural data management? These issues will be discussed in this chapter.

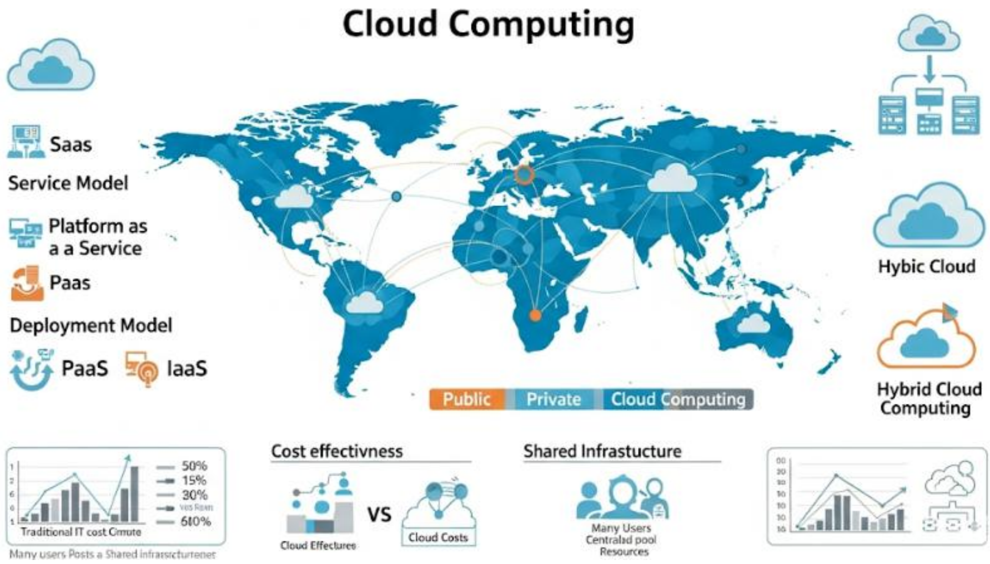

Fig 4 . 1 : Understanding Cloud Computing: Distributed Resources, Diverse Services

Conceptually, cloud computing consists of a distributed computing system composed by pools of physical resources, often distributed geographically that are managed by a software platform. Upon this infrastructure, a set of services that allow third parties and end users remotely access a diverse range of virtual resources and complex applications over the network are offered in exchange of fees, generally usage-based or subscription models. The term cloud computing comes from the metaphor of representing the Internet in the form of a cloud. Because cloud computing is a general term that refers to a combination of many technologies, two major taxonomies have been proposed to help understand its different aspects. The first taxonomy consists of service models that represent the type of services provided through cloud computing. These service models can be differentiated based on how many layers of the distributed computing architecture they cover. The second taxonomy represents how these distributed computing systems are implemented, and it is known as deployment models.

Although there is a number of hybrid models, there are three main pure deployment models. They are known as public clouds, private clouds, and hybrid clouds. Cloud computing is enabling a new way of providing IT services. Through a shared infrastructure and usage-based pricing, cloud computing allows customers access to the same high quality services that were previously available only to large corporations, but at a fraction of the price. Indeed, paying only for what you need, when you need it, allows for significant savings. Furthermore, the cloud customers shift the burden of service management and maintenance to the cloud service provider, which, as a specialist in providing that service, is far better equipped to manage the service efficiently and at lower cost.

4.2.1. Definition and Key Concepts

Cloud computing is a broad spectrum of IT services delivered over the Internet, serving multiple customers on-demand from a shared and scalable pool of resources. Cloud is one of the breakthrough technologies in the IT department introduced in Sixties. In cloud computing, infrastructure and software applications are accessed on demand via the Internet, rather than built and managed in-house. Cloud computing is based on the advanced concept of utility computing, agreeing on flexible and dynamic pricing, which operates only on demand. In the same way users access infrastructure such as electricity via the grid, they can access a broad range of IT services online as private cloud, public cloud, hybrid cloud and distributed cloud. The IT services could include data processing, storage, application hosting, software, security and network management.

Cloud computing concept eliminates the cost and complexity for organizations. Cloud computing adds increased flexibility to the solution and allows companies to leverage technology larger and the solutions that were not previously practical to them. With cloud computing concept, anyone can use a large scale solution for small money. Technologies such as virtualization, load balancing and increased bandwidth are taking the perfect base for implementing cloud computing solutions. The next generation of computing complexity will solve the technologies such as cloud computing, grid computing, and utility computing. Cloud computing is intended to provide computing power in a way that is similar to the electricity grid. Virtualization technology is being used to introduce these technologies. A computing cloud is a phrase often used to describe abstracts to resources in cloud computing concept technology. It is organized primarily because a remote cloud provider owns the resources.

4.2.2. Types of Cloud Services

Cloud services involve a collection of tools that allow users to manage activities such as storage, security, and functionalities without worrying about infrastructure. In the Public model, they need to pay for what they consume. Private models are dedicated and allow to make changes to the configuration. Furthermore, Community models involve several customers and an intermediate price. Hybrid models combine several Cloud scenarios. There are several types of the Cloud, differentiated according to their Service Models. The models differ in the level of abstraction that the service provider exposes the users. Three models concentrate the major demand in Cloud services: Infrastructure as a Service, Platform as a Service, and Software as a Service.

Infrastructure as a Service is the model in which the infrastructure is the service exposed by the provider. The clients have complete access to the infrastructure, having to manage the operating systems and the middleware. Infrastructure as a Service is mainly directed to users that need infrastructure to run arbitrary software, such as major organizations and research institutions. Software as a Service is characterized by the provider exposing the Software, that the users access from equipment with very small processing capacity. Software as a Service is mainly oriented to end-users, who need to run specific software with low processing and storage requirements. Platform as a Service model involves an intermediary solution. The provider makes available the infrastructure and middleware to the user, which has to manage the applications on the infrastructure to run specific functionalities. Platform as a Service has been presented as a good alternative for organizations that want to deploy specific services without investing in the complete development of the components. According to the respective level of abstraction, the models can be used interchangeably. Thus, a service offered through a Platform as a Service can be exposed via Software as a Service, and an Infrastructure as a Service can be exposed via Platform as a Service.

4.2.3. Benefits of Cloud Computing

The cloud concept is often seen as a delivery technology to provide computer resources. Using the cloud is a natural step for any company. However, the answer to the question of why it should be considered is more complex. The business benefits associated with cloud solutions are quite considerable. Increased business opportunity, reduced costs, and increased opportunity for strategic focus are some of the main business benefits. In addition to this, they extend the reach and power of IT through increased collaboration, increased innovation, business flexibility, and the ability to deal with risk and disaster recovery, all at reduced cost.

The use of the cloud can decrease many of the complexities and costs that companies have today with IT solutions. Companies can take advantage of economies of scale. Companies are increasingly taking a whole-ecosystem view of costs and benefits. The business solutions that operating in the cloud can make possible could be shared and generate revenues for their businesses. Businesses can allow IT to focus on strategic issues that align with the goals of the firm. Cloud computing services can be dispersed globally and lower costs at the same time these operations allow companies to consider requirements of closer locations for newer products and services that better cater to their specific market. Reduced cost models can be made available to consider maintaining less of the IT infrastructure and software internally within the firm.

Companies can create a model where investments in IT are predictable and can be directly tied to certain target metrics. There is a big effect associated with cloud computing and many of the benefits may be difficult to quantify. Still, the fact that they are becoming more common as firms adopt the new model demonstrates their importance.

4.3. Agricultural Data Management

Agriculture plays a significant role in the economy of any country, especially for developing countries. It is also one of the primary sectors of Indian Economy. Agriculture, being a major economic activity in the country, is confronted with greater demands on an international scale. In this scenario, India is presently not only self-sufficient in food production, but also exports food grains to other hungry nations. In order to achieve and sustain this notable goal of improving production and productivity, it is necessary to ensure better management of agricultural data and agriculture-related information, as natural and technical resources are highly sought for both agriculture and commerce. Decisions at the operational, tactical, or strategic level with regard to agriculture have several dimensions and are highly complex in nature due to risk and uncertainty factors, multiplicity of objectives, and the inherent social and economic milieu.

Diverse information is needed in the various areas of agriculture sector, which can be classified into various types as statistical data on area, production, and productivity of crops; information on investment in agriculture and productivity growth; information on rural wages and prices of agricultural inputs and outputs; information on employment in agriculture and its relation with economic growth; information on important sectors of rural and agricultural development; information on problems and prospects related to trade in agricultural commodities like cereals, plantation crops, oilseeds and edible oils, fertilizer, agro-chemicals; information on technology; information on credit requirement; information on policy issues; and general information on rural infrastructure, resources, and services.

4.3.1. Importance of Data in Agriculture

Agriculture has played a significant and vital role in human society and civilization from time immemorial. Humans transitioned from a nomadic way of life to a systematic settled life with the advent of agriculture. Human beings underwent gradual growth and development during the process of evolution. The procedure of obtaining our daily food from nature was first discovered by the ancient people, ultimately leading to the establishment of agriculture. Agriculture is one of the oldest and the most significant core areas of human life. Although the early methods of farming did not require high levels of expertise and latest technology, it was the first step taken by human beings in the direction of progress. The primary idea behind all sorts of agriculture is the technology-based production of food, fuel, fiber, and other goods and related services by the systematic and controlled use of living organisms. With the advancement in technology, farming has modernized into a more regulated and supervised way of living. Progress in agriculture is not only the progress in food or flower production and marketing; it also involves applying agricultural science in agriculture such as genetic improvements and those enhancements that help increase accessibility to food. Like any business or organization, progress in agriculture requires reliable data and information for making decisions. Information provides knowledge, which is the most important resource for smart decisions. Implementation of latest and upgraded technologies, while necessary, may not be sufficient. Knowledge is required to interpret, choose, and adopt new technology. Agriculture often involves increasingly complex decisions needed to compete and survive; for example, advancing technology is changing production procedures. Information helps reduce uncertainty. Decision making and uncertainty reduction require more information for quantitative analysis than in previous years. The decision-making problems of primary importance require more and better information.

4.3.2. Types of Agricultural Data

The vast and increasing volume of agricultural data available has contributed to the demand for cloud-enabled technologies. Unfortunately, few traditional data management tools are able to successfully manage this data due to the specialized and complex nature of many data types. For example, submissions of field transportation data may consist of thousands of fields representing the start and completion timestamps; the GPS coordinates; and the weight of the starting and completed loads for each trip. Accumulations of transportation data for a large number of fields, hauling season, and the corresponding transporters can produce very large datasets. Various data collections related to individual fields, including equivalent software files, can be collected from agricultural equipment and sensor technologies. These collections of field data could potentially number in the millions yearly, each encompassing multiyear histories of agricultural equipment use and agricultural practices. GPS data from equipment motion analyses may be important for monitoring purposes, but the requirements for sensor collection strategies can create significant challenges.

Field weather data, ranging from moisture sensors to UAV-collected high-resolution multispectral and thermal images, may number in the millions from agricultural activities year-round. Satellite remote sensing technologies have also matured for use in agriculture. The long data histories and new sensor data streams could be used to develop sophisticated services for use by farmers and governmental agencies. These data services could assist in farm management recommendations, hazard monitoring, and support the development of precision ag services. Crop yield maps, which were enabled by GPS and yield monitors on harvest equipment, can provide important multiyear analyses about a field's reliability and variability of production. However, the traditional methods of generation and use of crop yield maps could create significant impediments to their timely use.

4.3.3. Challenges in Data Management

A major challenge in agricultural data management is the growing volume of data being generated, collected, and stored. The rapid evolution of cloud computing, sensing, and networking technologies allows agriculture to adopt the rapidly-growing Internet of Things that is already prevalent in other sectors such as smart, connected homes and cities. IoT-enhanced agriculture will lead to exponential growth in the types, volume, and velocity of Big Data generated. The volume and complexity of the data created will certainly exceed the capabilities of traditional data management systems to ingest, filter, analyze, and make informed decisions in real time, particularly for the critical and dynamic near-farm and on-farm applications.

Another key challenge is the disparity in the formats, structures, and protocols of the collected data. Agricultural data is collected from various data generation and collection sources such as remote sensing images, GPS-ingested pontoons, on-board data collection systems, manually reported data points, etc. The sensors used in different systems or devices may not be standardized. For instance, the information provided by two different companies for local weather data may have different granularity; one may provide hourly forecasts while the other daily. The type of information being analyzed may be different; one may analyze matter soil temperature while the other wet soil temperature. The frequency of sampling temperature data may differ. Ignoring these protocol mismatches could result in incorrect decisions being made. Another major consideration is that the organizations that own these data sources may either physically store these data in geographical locations where data access would be very costly, or imposing data access fees that would be very high in comparison with the few other sources of such data in the region.

4.4. Scalable Solutions in Cloud Computing

1. Elasticity and Scalability

Elasticity is the scalability feature in cloud computing. It is the ability of a cloud service to quickly provision, deploy and transfer critical workload resources in and out of the cloud. Cloud computing provides easy, on-demand, variable provisioning of resources.

Several tasks in life sciences are demanding with heavy resource utilization. This fact in combination with increasing speed of the cloud access motivates researchers and practitioners to diagnose, optimize, and perform these tasks some of the times in the cloud. A classical example is the daily workload in many hospitals. During workdays demand peaks in usual working hours while the available computing capacity diminishes to a minimum overnight. Using a cloud computing service permits flexibility in task execution. When operations in the cloud are not internal and can be accessed externally requesting cloud resources from a cloud provider adds more convenience and flexibility. Cloud providers allow hiring resources with a very short notice. Tasks such as DNA sequence assembly, genomics application or particle simulations must be carried out with very intensive resource usage. One option is to hire clusters of compute nodes during high demand. Another option is a resource combination of core nodes which is kept for tasks demanding very sequential CPU transfer followed by distributed core tasks making heavy use of data transfer.

2. Cost-Effectiveness

Unpredictable and sustaining demand peaks on computing resources are the main motivation for scientists and businesses to use cloud computing. The premise of the cloud paradigm is the trade off of reliability, availability, and performance of the on-being-tasked tasks done in cloud for outsourcing the on-serving cloud resources. Well-provisioned data transfers between cloud customer and cloud services ensure cost-effectiveness by shortening transfer delays as well decreasing monetary costs for data transfers and cloud service usage. Data transfers in the cloud environment have a high potential for optimization of the transfer speeds and transfers costs. Task optimal provisioning in using sensor networks, wireless wide area network, and cloud resources ensure the highest savings.

4.4.1. Elasticity and Scalability

The Cloud deployment is focused on the exposed services model with high elasticity and scalability. The high elasticity model means that most of the times semantic services are being posed, as opposed to raw computing services on top of the raw configurations. A service might be called on-demand or regular on-demand, and the large user base that is different by application groupings and sizes will contribute to increase the amount of services on job resubmission, Life Cycle management, Development Systems and so forth.

However, semantic services by user groups are finite and the actual increase in number of semantic services cannot increase proportionally to the amount of the users. A user call to a service is bound to the availability and the scalability of the Cloud Resource

Pool. As the amount of resources will increase when the pool gets utilized, more and more user groups will be able to access the elastic capabilities of the exposed services. It is the combination of the passing by requests that will be actually enabling the Cloud potential elasticity. As opposed to raw computing services model, who have an unlimited scalability as job queuing increase.

Service Chain Scalability is another Cloud unique concept. The generic Idea is that any Service can leverage Core Service Capabilities and Scale without limit. These Core Services can be provided by the Cloud Resource Pool Provider and scaled like the others. Not only the Infrastructure but scalable services expose a unique capability to account for the resource that are really needed in a pay model. Service Chain Scalability has to do with Birth to Death Map Count for Service Enablements.

4.4.2. Cost-Effectiveness

Apart from the resource elasticity and scalability of cloud platforms, cost-effectiveness is another critical reason for why they are the preferred choice in implementing scalable solutions for data management in multiple application domains.

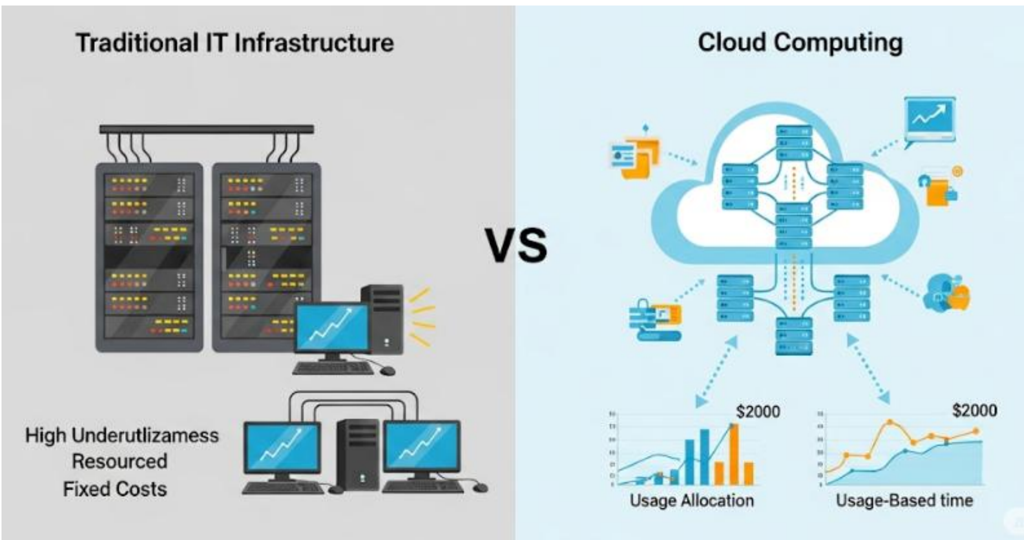

Fig 4 . 2 : The Economic Advantages of Cloud Computing

Although investors have been motivated to build large data centers of traditional IT infrastructure due to growing demand for resources for both enterprise and business, they want to maximize the utilization of the infrastructure in terms of cost. Unfortunately, in traditional IT, processes have a hard and fixed service agreement with a particular hardware and other infrastructure resources which are not shared, limiting the flexibility to make workload based sharing of data center resources. Cloud makes it easy for both the stakeholders and the consumers to only pay for resources when needed. For example, consider the case when a user needs to perform a small burst of processing reminder task on all the input reminders in a few minutes. In traditional IT, if the company needs to perform that small task every so many days, the processing logic reminders have to be present in the core IT architecture which will lead to idle time for most of other days or hours. But with a cloud approach, the cost will be directly related to the time spent and the minor additional infrastructure required to handle that particular task. Although this small example does not consider any overheads in transferring data to cloud and getting the results back, it still demonstrates the cost effectiveness of cloud computing infrastructure at a micro level.

4.4.3. Performance Optimization

Optimization is a technique that allows both users and providers to enhance cloud service capabilities or resource utilization depending on unique requirements. While optimization can occur at several levels of data management and computing services, we focus on two key components that may improve performance, namely cloud service optimization and data access optimization.

Most users of cloud computing services automatically assume and rely on service and performance guarantees provided at different service levels. These include specific levels of availability and fault tolerance of storage and hosting services, response times for payment transactions, and query response times. In principle, these are all operations that providers must handle transparently for customers with no adverse consequences on the service user experience. Customers expect to receive the performance levels that they pay for regardless of demand pressure on providers. At the same time, progressive cloud service providers who aggregate and serve large groups of customers via common physical infrastructures are continuously optimizing their own services and internal processes, typically by innovating in three key areas.

Cloud providers work with increasingly high levels of resource virtualization using well-developed tools and methods in partnerships. Within the cloud computing ecosystem are a number of developing optimization platforms to integrate cloud services to enterprise management tools, typically to seamlessly address shifting demand or predictable long-term usage cycles. Some enable enterprise clients to avoid peak costs by better monitoring usage while taking advantage of discount pricing plans.

4.5. Implementing Cloud Solutions

Various Cloud resources are being constantly updated. Therefore, it is appropriate to learn about Cloud based solutions by implementing them. In our experience, data and applications should be migrated to the Cloud in a phased approach. ERP solutions can be used to manage each step of Migration for the organization. Scalability should first be implemented for a small sample of high priority data and applications serving a small sample of users. Both server and client sides need to be prepared for such a process. Later, those components can be scaled to do more tasks with more data and serve more users. Eventually, the data and applications of the organization can be entirely migrated. This phased approach could lead to learning difficulties and lacking organizational memory of the decisions taken during the earlier phases. However, many benefits can accrue to having a phased approach in terms of experienced learning over time and an evaluation and gauging process to analyze the impact of Cloud-based solutions. Dedicated teams need to be created to carry out the Migration process. Selecting a Cloud provider is a critical choice with high risk degrees. Such a Cloud provider can usually provide a staff member that works closely with the Migration teams.

The choices of Cloud capabilities and providers are regularly updated. We need to look for suitable capabilities for our needs. Some examples of Cloud capabilities of use for the described Migration steps are multi-tenancy, data sharing and software distribution services, messaging services, VM images for different processing platforms including mobile ones, parallel computation platforms, workflow management services, long term storage solutions, data cleansing and transformation tools and services, backup and disaster recovery services, security management services, virtualization and containerization tools and services, remote repository services for code, data, documentation and toolkits, debugging and optimization tools and services, accounting and cost management services, user management and administration services, application and API monitoring and performance management services, business rules management services.

4.5.1. Planning and Strategy

Agricultural data management is essential in the rapidly developing smart farming area. With the exponential growth in the scale and complexity of the corresponding processing and analysis workflows, bigger computing capacity is needed to meet the data processing demands. In recent years, cloud computing has become the de facto standard to answer the higher computing demands from various big data applications in various areas. We discuss the challenges, decisions, and solutions, derived from our years of field experience designing, developing, and deploying public cloud solutions in addition to on-premise high capacity and high availability enterprise system solutions for agricultural data management.

The key challenges with the public cloud solutions are the design and management of the cloud systems to ensure desired Quality of Service attributes and the associated total cost of ownership. Cloud computing has matured enough to offer ready-to-use systems and solutions for upper layers of the data processing pipeline, which provide convenience in their usage. Even then there are several key challenges needed to be followed when using public clouds for agricultural big data solutions, compared to using them for other applications. These are mainly due to the unique characteristics of the agricultural workload patterns. These types of challenges are, how to decide what type of data should reside on-premise and which parts should be deployed in public clouds. How to take the right decisions properly in benchmarking, selecting, and sizing diverse cloud services with minimized costs. How to design intelligent solutions with appropriate cloud bursting strategy to minimize cloud costs, optimize responsiveness, and ensure the availability required by the agricultural system.

4.5.2. Selecting Cloud Providers

Hybrid clouds are combinations of public and private infrastructure capabilities designed and implemented in such a way that services and workloads can move between the public and private clouds as needed. Due to the transient nature of many agriculture workloads, especially those with very variable and unpredictable workloads and the need for low-cost capabilities, hybrids can be very useful. Conversely, some cloud workloads may need to be processed on site, such as for data security and privacy reasons. These workloads would be performed on private clouds where the data remain inside the firewall. These private clouds can be designed with capabilities that allow for similar economies of scale as hybrid clouds.

There are many private and hybrid cloud providers in the agriculture space. These include large providers who have roots in both the private cloud space and the hybrid cloud space and are in the process of building agriculture partnerships to use their technology offerings. They view agriculture as one of the key growth markets. There are also a number of larger commodity vendors that have been increasing their hybrid cloud capabilities due to their use in manufacturing activities involved in food processing. However, even the smaller vendors with their own cloud offerings, especially those with strong IoT offerings based on tight alliances with devices and sensor providers should be considered for partnerships as they may be able to provide lower cost solutions that are designed specifically for agriculture. Such partnerships might make the timing and rewiring of cloud computational activities easier for smaller players.

4.5.3. Migration Processes

Migration is a process of moving data from one storage or computing system to another. This is done to support more sources, more load and to provide redundant backups or failover. The practice of migrating agricultural data from its previously used desktop/laptop computing environment to a distributed cloud-based system for agriculture is thus not new, but it is often one of the first hurdles that the user needs to address and is thus valuable to address as one of the first steps in the implementation process. There are two steps, or types of transfer, usually involved in migration. The first is an initial bulk data upload or transfer that must be performed, during which the previous storage and computing environment need to be synchronized with that available in the newly setup cloud system to enable the processes and tools in the new cloud environment to use the data. The second is the subsequent incremental update process that is used to send any updates performed to the original dataset in the previous stored environment since the time of the first bulk upload. The incremental transfer process is generally of smaller size compared to the initial transfer and can thus be performed more frequently to keep the datasets synchronized. The initial bulk data transfer may be performed by a direct connection or network upload of the dataset from the previous environment to the cloud environment. Some cloud providers may also provide other external upload methods, allowing for the bulk initial transfer to be performed by sending a hard disk copy of the source dataset for them to then upload into the cloud environment on the actual cloud storage service.

4.6. Data Security and Compliance

Data security and compliance are paramount needs for cloud computing solutions in the agricultural sector. Sensitive information is sought by malicious third parties and is at risk of unauthorized use and hacking. Information security must be protected, which includes technology security, natural security, physical security, collaborative security, managed security, compliance, and management plan control. Data must be sensitive and private for hidden terms to be withheld, monitored, tracked, and reported to the rightful employer or owner. Security threats are real and present. The data recorded is often unrecoverable; reputations can be compromised with heavy losses attached. Questions regarding who owns the data stored in the cloud arise. Data is often encrypted and finally managed by the cloud provider who holds the unique key to that data, thus retaining ownership. Cloud data also runs the risk of access by insiders, given that their business relies heavily on the data stored in the electronic cloud.

Fulfilling regulations, laws, and compliance requirements is essential with respect to the jurisdiction of the entity that governs the cloud provider that holds the cloud data. These vary from one country to another and are heavy in the privacy space. Data security goes beyond technology either with respect to the cloud provider or the entity accessing the cloud-based infrastructure. Many tools are available to assist with this responsibility. Containers and virtualization are well established technologies that allow the needed flexibility, resource saving, and device software technology access. These two technologies combined together with compliance requirements, industry guidelines, and security frameworks that assist cloud providers and users will help any organization become compliant. Combining these security frameworks will expedite an organization's compliance with these security requirements.

4.6.1. Security Challenges in Cloud Computing

Cloud computing offers a compelling alternative to traditional information technology services by providing flexible, efficient, and scalable access to storage and processing resources. Modelled on utility services such as electricity, computing resources can be quickly provisioned in any quantity and accessed over the Internet using pay-per-use business models. Users of IT become free to focus on their core businesses as cloud services relieve them of the huge upfront capital expenditures and ongoing personnel and technological demands of maintaining dedicated IT resources. For major corporations, cloud computing promises a reduced strategic focus on IT services and potentially reduced costs. For small and midsize businesses, cloud computing makes it possible to operate with the same high quality and level of service as the multinationals.

Cloud computing's rich set of services and flexible business intelligence tools provide capabilities to support prediction analysis and identification of marketing channels, but the promise of cloud computing comes with a new set of concerns for IT, including: a volatile technology infrastructure space; potential vendor lock-in and loss of long-range strategic focus; potential performance overhead; potential data security violation; and potential legislative compliance violation. Cloud computing is a new multi-user version of cyberinfrastructure. Like the Internet, the cloud is a distributed, decentralized user device model with large-scale, long-range routed applications, where the cloud provides its users with events and messages from both inside and outside of the cloud service. In addition to traditional IT security challenges, such as secure virtual machine accesses, secure authentication and authorization procedures for cloud service usage, secure configuration structures, and secure data communications, cloud computing presents its user community with an additional set of security challenges. These challenges are based on the types of actions that incur a lost trust of their cloud service providers.

4.6.2. Regulatory Compliance

As industries all over the world have shifted to the cloud for ease of operations, reduced costs, and better services, laws and policies have evolved to cover data storage and processing violations that lead to compromises of data security and privacy. With the implementation of various regulations, cloud data operators need to be diligent in following the rules in every jurisdiction relevant to their businesses.

Farmers and agribusinesses across the globe are being pushed by market pressure to adapt to data-driven innovations in agricultural solutions across the agri-food value chain. In doing so they are bringing their data to the cloud, where it is unencrypted and accessible to third parties and foreign actors that federal rules may not restrict. The information might include data about daily farm operations, sensor data, and imagery data from UAV flights, especially around harvest, planting, and other critical operational activities. However, laws and industry guidelines are still catching up to address issues around third-party access to data that are undeterred by any contractual restrictions on its use.

The reasons to remain aware and concerned about cloud services and their ability to expose data broadly through international channels are twofold. The first is due to issues of data ownership, security, fair use, and intellectual property on the part of the data subject or the individual and the second concern harkens back to the issue of data protection from all the nuances and associations linked to chronicling business decisions and patterns.

4.6.3. Best Practices for Data Security

Data security is always a concern when sensitive data is involved, and the highly regulated field of agriculture has an additional responsibility of protecting PII. As a result, users, data owners, and cloud providers need to follow best practices that are essential for maintaining data protection, privacy, integrity, and availability. Security in and of itself can be achieved through various means based upon user needs as well as cloud provider resources. While it is impossible to ever completely remove the risk of a major data theft or unauthorized access to PII, following certain best practices can significantly lower the risk.

These best practices can be divided into User Best Practices and Data Owner/Provider Best Practices. For the user of cloud resources, access management is crucial to ensure that only authorized personnel have access to systems where PII is implemented or stored. Enabling multi-factor authentication prevents unauthorized individuals from gaining easy access to sensitive information. The responsibility of user account management, including resource usage monitoring and regular password changes, falls on the client-side managing networked devices and cloud resources. Encryption of all sensitive data before uploading to cloud resources will ensure that data cannot be easily decrypted regardless of successful data theft. Additionally, conducting regular security audits in order to inform cloud resources management as needed will only strengthen security.

Data owners and cloud providers can ensure the protection of user data through backend implements. In addition, implementing the principle of least privileged access and Data Loss Prevention policies ensures that only authorized users view sensitive data. Despite encryption during transit and throughout processing, which is a large security focus of most cloud providers, PII can still be at risk of being accessed or viewed by unauthorized parties if it is ever stored unencrypted.

4.7. Future Trends in Cloud Computing for Agriculture

Cloud computing has transformed the agriculture landscape into one where farmers can work more productively and strengthen food security. Enabling technologies such as big data analytics, IoT, machine learning, and AI aggregate large data sets, analyze these data, and automate agriculture processes. Consequently, the benefits of cloud computing are far-reaching, deploying solutions that enable precision agriculture to increase crop yields, and responding to pivotal challenges including pest and disease control and improved access to agriculture forecasts. What does the future hold? Inevitably, the transition of cloud computing technology into both general and domain-specific solutions will depend in part on the envelopment of new technologies, and on vertical sectorisation. Emerging technological trends include the development of superfast processing speeds, storage capabilities, and network transmission speeds at low costs. Partially these development are driven by the exponential increase in data generation, which have been termed the “data explosion". By 2025, the cumulative sum of previously generated data globally will reach 175 zettabytes. Data storage globally is at 4.5 zettabytes in 2020 and will reach 22.4 zettabytes by 2025. Hence, the growth is by nearly 500%. As more people gain access to cloud services for data storage and processing, cloud data traffic on service providers, including multicloud application and service providers will have more infrastructure increased to offer. These increases in infrastructure capacity will also be accompanied with the proliferation of more innovative cloud solutions that meet the demands of both consumers and enterprises. Vertical sectorization will occur because of both sustainability and data ownership concerns. As emerging technologies facilitate a lower cost of service delivery, demand for solutions that meet both sustainability standards and accommodate organizational data ownership concerns, such as on-premise private data centers, or hybrid data delivery solutions, is projected to grow amongst organizations that have a significant data security profile, including those in sectors such as finance and healthcare. However, farmers in developing economies must also benefit from lower cost domain-specific solutions that have been proven to promote and facilitate sustainable agriculture practices. The digital divide must be overcome.

4.7.1. Emerging Technologies

Cloud computing has been a pivotal technological intervention that has exponentially changed the way agricultural IT solutions are designed and typically made available to end users. Alternatives to conventional centralized models provide to researchers, research institutions/agencies, and companies engaged in any aspect of agricultural data usage the ability to scale up their services. However, as can be expected, any paradigm shift is accompanied by a huge gap between deployed technology and the highest potential it can deliver. There are emerging technologies in all the generic disciplines involving various aspects of cloud computing. It is very likely that many of these will be increasingly utilized for agricultural data management, providing unprecedented new opportunities.

Machine learning has changed the way data processing systems are being implemented. Its essential features of scalability, ease of implementation, and, sometimes, the ability to mishandle data imperfections and still provide suitable solutions, have popularized the use of machine learning for many application areas, most visible in vision data processing areas like self-driving cars, automated detection of human emotions, surveillance, and similar applications. Cloud architecture provides the necessary horizontal and vertical scalability for implementing deep learning-based machine learning algorithms successfully in a short time. Further, enabling increasing levels of automation of image processing and increasing democratization of these technologies are bringing visualization technology to the world of non-experts, allowing users without machine learning skills to apply these powerful capabilities to extract information from images and videos. The agricultural sector generates vast amounts of data in the form of images and videos and has largely untapped the latent potential of machine learning capabilities of clouds to provide clear, interpretable, and actionable outputs from such data.

4.7.2. Sustainability and Cloud Solutions

Cloud computing is playing an important role in increasing the performance of various areas and shows a prominent impact in many others. Sustainability and sustainable development have become an important issue in every area. Data needed to point toward a sustainable solution are generated in all areas without exception, but many overlook the opportunity for utilizing technology to empower those areas. Researching on the solutions for sustainable agricultural development needs comprehensive data management of all areas regarding various infrastructures. Cloud solutions applied to any other area are supporting management in terms of lowering the operational costs and energy consumed on physical infrastructures. Agriculture is the largest consumer of worldwide freshwater resources, and one of the significant challenges of this century is to increase the productivity of freshwater for diverse conceivable uses. Agriculture is considered to be a large source of carbon emissions, which impact climate change and intensification of the effects of climate change on agriculture, and has an effect on food insecurity across the globe. Many of these challenges can be addressed using new ICT applications based on IoT, cloud, and big data technologies providing the technological infrastructure for the application of sensor and satellite-based remote sensing.

When such technologies will be combined with appropriate policy and institutional arrangements, they will largely contribute to achieving zero hunger – another of the Sustainable Development Goals. The upcoming decade is critical for accelerating the deployment of innovative and optimized services for natural hazard monitoring, including activities for detecting changes and addressing specific climate-related challenges in sustainable agriculture, forestry, and land use. The dynamic models that will be developed will be driven by remote sensing data, but also field data, integrated and managed in the cloud with cloud services that will allow sharing, processing, and selectively accessing the large amount of data generated.

4.8. Conclusion

The rapid growth of agricultural data management in recent years clarifies that cloud computing infrastructures are becoming privileged tools for developing large-scale computing solutions in different agricultural disciplines. Firstly, we provided an overview of the main cloud computing enterprise solutions available on the market, highlighting that one platform is potentially the most suitable for agricultural research. Then, we described how to improve the development and the integration of services to manage agricultural data with efficient and secure authorizations and big data storage and processing. Finally, we presented case studies to demonstrate how to use the solutions proposed as benchmarks for other agricultural disciplines or research.

Precisely, this platform offers some advantages as solutions for users. Collaboration with organizations can greatly increase the collective impact of agriculture research. Indeed, data curation is an essential step to guarantee reproducibility of analytics. The elaboration and curation phases of research data in the proposed pipeline can take advantage of most of the services, such functions into adjustable cloud-based containers, guarantee scalability while resulting easy to integrate with both each other and other third-party services. In addition, the combination of training technologies coupled to trying configurations that are simple to set creates a favorable environment for new practitioners, mainly because of the low level of complexity on both the user and service-use-expert sides. The detection of complex patterns in research data with higher accuracy levels might become a starting point for the automation of several productive cycles also in developing countries where labor cost is usually cheaper than in developed countries.

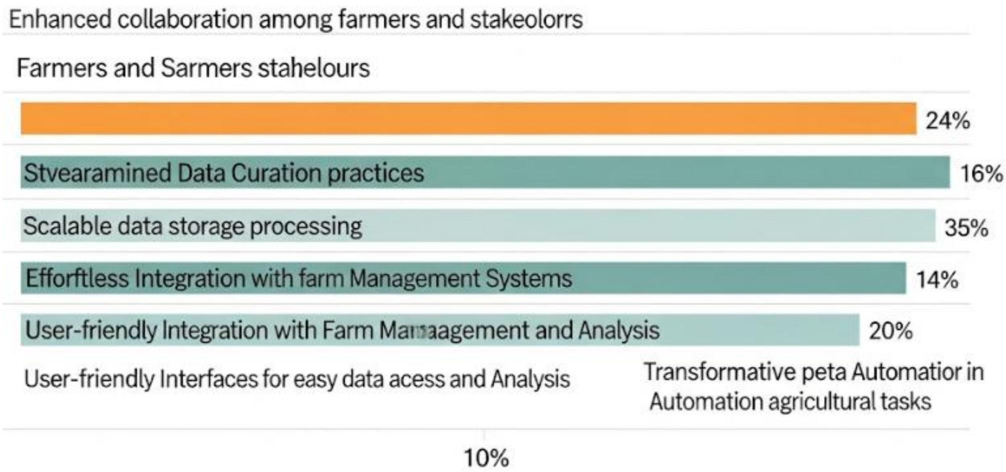

Fig 4 . 3 : Powering the Future of Agriculture

All the contributions expressed in this paper can be seen as a first attempt to create the basis for a global computing platform able to host reliable and user-friendly services for big data management in agriculture. In conclusion, the development of reliable scalable cloud solutions that are easy to use may represent a tremendous opportunity to support automation in agriculture. A potential return on investment would surely encourage partners to support the creation of such infrastructures.

4.8.1. Summary of Key Findings and Implications

In this paper, we explore the collective potential of cloud computing and cyberinfrastructure to spur the next Green Revolution by enhancing the efficiency and efficacy of both basic and applied research in the agricultural sciences and, in particular, precision agriculture. For several key areas of the ongoing agricultural research questions, our analysis suggests that current architectural solutions, based primarily on local data storage and controls, technologies, and ground-based sensor networks operated by small teams of local researchers and farmers, are neither sufficiently scalable nor cost-effective. There are, however, existing and emerging solutions grounded in cloud and cyberinfrastructure technologies that could address many of these constraints.

Enabling technologies include ultra-scalable networking technologies; remote sensors and geospatial-analysis technologies; more effective resource allocation and predictive models for damage; novel wireless networking schemes that deliver higher bandwidth at lower cost; remote sensors and geospatial-analysis technologies; improved methods for analyzing and visualizing massive data files; novel assessment, early-warning, and remote-control technologies; enhanced airborne or satellite-based remote sensing technologies; and greater deployment and integration of the appropriate sensors in specific crops.

Recent early moves into technology-augmented agriculture, while largely based on the Internet and local network solutions, could also benefit from a stronger cloud and cyberinfrastructure footing. For example, new venture farm and big data monitoring programs are currently limited to localized networks. While entertaining, the ambitious concept of a global sensor network, operated and serviced by small viable settlement satellites and drones, still raises technical issues. Most critical is bringing specific satellites' geosynchronous orbits closer to the ground while maintaining constant coverage of specific areas and bringing costs down far enough to allow for reasonable fees for small landowners.

References

Wolfert S., Ge L., Verdouw C., Bogaardt M.J. (2017). Big Data in Smart Farming – A Review. Agricultural Systems, 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Jayaraman P.P., Yavari A., Georgakopoulos D., Morshed A., Zaslavsky A. (2016). Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt. Sensors, 16(11), 1884. https://doi.org/10.3390/s16111884

Kamilaris A., Gao F., Prenafeta-Boldú F.X., Ali M.I. (2016). Agri-IoT: A Semantic Framework for Internet of Things-Enabled Smart Farming Applications. IEEE 3rd World Forum on Internet of Things (WF-IoT), 442–447. https://doi.org/10.1109/WF-IoT.2016.7845471

Brewster C., Roussaki I., Kalatzis N., Doolin K., Ellis K. (2017). IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot. IEEE Communications Magazine, 55(9), 26–33. https://doi.org/10.1109/MCOM.2017.1600528

Sundmaeker H., Verdouw C.N., Wolfert J., Perez Freire L. (2016). Internet of Food and Farm 2020: A European Use Case Approach for Smart Farming. Sensors, 16(9), 1461. https://doi.org/10.3390/s16091461