


Chapter 5

Data Engineering for Smart Farming Building Scalable Agri–Data Pipelines

Rakesh Ramakrishna Pai

 <https://orcid.org/0009-0007-0129-3633>

Independent Researcher, USA

Venkata Manikesh Iruku

 <https://orcid.org/0009-0007-9109-7598>

Independent Researcher, USA

ABSTRACT

Smart farming is revolutionizing agriculture by leveraging data-driven insights to optimize crop yields, resource usage, and operational efficiency. However, the success of smart farming initiatives heavily depends on the design and implementation of scalable, robust agri-data pipelines. This chapter explores the critical role of data engineering in modern agriculture, focusing on techniques for collecting, processing, and managing vast volumes of heterogeneous agricultural data. We discuss the integration of IoT sensors, drones, satellite imagery, and weather stations, along with methods to handle challenges like data heterogeneity, real-time ingestion, data quality, and storage scalability. The chapter also presents architectural patterns for building end-to-end agri-data platforms, including the use of cloud-native solutions, edge computing, and AI-driven analytics pipelines.

DOI: 10.4018/979-8-3373-5283-1.ch005

INTRODUCTION

Agriculture, one of the oldest and most vital human activities, is undergoing a profound transformation. Driven by technological advancements, especially in data science and engineering, the agricultural sector is evolving into what is now widely termed as *smart farming* or *precision agriculture*. This transformation is essential to meet the increasing global demand for food, optimize resource usage, adapt to climate change, and ensure sustainability. Central to this revolution is the role of data—data generated from farms, sensors, satellites, weather stations, and even from farmers themselves. However, the mere availability of data is not enough. It is the efficient collection, processing, analysis, and management of agricultural data that truly unlocks value. This requires building scalable, resilient, and intelligent agri-data pipelines—a domain where data engineering becomes pivotal.

The traditional agricultural practices, largely dependent on manual observation and experience-based decision-making, are no longer sufficient in a world challenged by rapid population growth, shrinking arable land, and unpredictable climate patterns. Modern agriculture seeks to integrate information technology with farm operations, resulting in the concept of smart farming. In this model, various digital tools such as Internet of Things (IoT) devices, drones, automated machinery, satellite imagery, and big data analytics are employed to gather actionable insights. These technologies promise improved crop yields, better pest and disease management, efficient water usage, and optimal application of fertilizers and pesticides.

Yet, the true backbone of smart farming lies beneath these flashy technologies—in the architecture that captures, moves, stores, and processes the data they generate. Data engineering for smart farming encompasses the end-to-end lifecycle of agricultural data, from real-time collection at the edge to deep storage in the cloud and subsequent analysis using advanced algorithms. Building effective agri-data pipelines requires addressing numerous challenges: data volume, velocity, variety, veracity, and value—the famous five Vs of big data—all apply acutely in the farming domain.

Firstly, the **volume** of data is enormous. A single smart farm equipped with IoT sensors can generate terabytes of data in a growing season. High-resolution drone and satellite imagery contribute even more. Secondly, **velocity** is crucial, as many agricultural decisions—such as irrigation control or pest management—need real-time or near-real-time responses. **Variety** adds complexity, with data coming in structured formats (like soil sensor readings), semi-structured formats (like drone images with metadata), and unstructured formats (like textual weather reports or farmers' notes). Furthermore, **veracity**—or the reliability of data—is a significant issue, as sensor malfunctions, environmental interference, and inconsistent data entry can lead to errors. Finally, extracting **value** from this data demands efficient engineering and intelligent analytical models.

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/data-engineering-for-smart-farming-building-scalable-agri-data-pipelines/383744

Related Content

Cruising Deep Convolution Neural Networks (DCNN) and Internet of Medical Things (IoMT) Fostering Smart Healthcare: Runway Towards Quality of Life and Wellbeing

Bhupinder Singh (2025). *Responsible AI for Digital Health and Medical Analytics* (pp. 211-240).

www.irma-international.org/chapter/cruising-deep-convolution-neural-networks-dcnn-and-internet-of-medical-things-iomt-fostering-smart-healthcare/365993

A Systematic Review of Gemini and Mathematics Education

Ali Maz (2027). *Encyclopedia of Modern Artificial Intelligence* (pp. 1-24).

www.irma-international.org/chapter/a-systematic-review-of-gemini-and-mathematics-education/407445

Principles and Methods of Explainable Artificial Intelligence in Healthcare: Framework for Classifying Alzheimer's Disease Using Machine Learning

Manimekalai K. and Abirami D. (2022). *Principles and Methods of Explainable Artificial Intelligence in Healthcare* (pp. 272-292).

www.irma-international.org/chapter/principles-and-methods-of-explainable-artificial-intelligence-in-healthcare/304185

A Methodology for Ontology Reuse: The Case of the Abdominal Ultrasound Ontology

Nur Zareen Zulkarnain and Farid Meziane (2019). *International Journal of Intelligent Information Technologies* (pp. 1-21).

www.irma-international.org/article/a-methodology-for-ontology-reuse/237963

Designing Scalable Location Based Games that Encourage Emergent Behaviour

Kate Lund, Mark Lochrie and Paul Coulton (2012). *International Journal of Ambient Computing and Intelligence* (pp. 1-20).

www.irma-international.org/article/designing-scalable-location-based-games/74367