Chapter 20
Sustainable Data Engineering:
Building Business Success With Eco-Friendly Innovations

Somnath Banerjee
https://orcid.org/0009-0006-5781-9051
AMFAM, USA

ABSTRACT

This chapter explores the integration of sustainable practices into data engineering to drive business success while reducing environmental impact. It examines strategies such as energy-efficient data pipelines, green data storage, and low-carbon cloud architectures that enhance operational efficiency and support sustainability goals. The chapter also highlights real-world case studies to illustrate the application of eco-friendly data practices in various industries. Readers will gain actionable insights into how sustainable data engineering can optimize resources, cut costs, and align businesses with global environmental standards, fostering both economic and ecological benefits.

INTRODUCTION

As the world becomes increasingly data-driven, the environmental impact of managing vast amounts of information is often overlooked. Data engineering, the backbone of modern business intelligence, requires substantial computing resources, leading to significant energy consumption and environmental strain. This chapter introduces the concept of sustainable data engineering, emphasizing the critical need to balance technological advancement with ecological responsibility.

The Need for Sustainability in Data Practices

The rapid expansion of digital infrastructure has resulted in a corresponding increase in energy use, greenhouse gas emissions, and electronic waste. Data centers alone are responsible for approximately 1% of global electricity consumption and contribute to rising carbon emissions. As businesses become more reliant on data for decision-making and operations, sustainable practices in data engineering are essential to mitigate these environmental impacts. The shift toward sustainability is not just a moral imperative but also a strategic advantage, as customers and regulators increasingly demand environmentally

DOI: 10.4018/979-8-3693-9750-3.ch020

conscious operations. The growing need for sustainability in data engineering has sparked a wealth of research focusing on energy-efficient practices and environmentally responsible technologies. Adnan and Mehmood (2021) emphasize energy-efficient approaches within data engineering to reduce environmental impacts. Similarly, Ahmed and Iqbal (2019) explore cloud computing as a key strategy for minimizing data center energy consumption. Al-Emran and Shaalan (2020) discuss the role of green technologies in enhancing sustainability in data engineering, while Anderson and Zhang (2022) address techniques and trends in optimizing data storage for better energy conservation. In line with this, Bhatia and Gupta (2020) investigate how data engineering can contribute to reducing carbon footprints, specifically in the context of green energy. Brown and Miller (2019) highlight eco-friendly cloud computing practices, demonstrating how cloud solutions can reduce environmental footprints. Chen and Xu (2018) delve into the development of energy-efficient algorithms for big data analytics, crucial for reducing computational energy usage. Das and Kumar (2021) focus on minimizing the environmental impact of cloud-based data pipelines, suggesting methods to improve their sustainability. DiPietro and Smith (2017) explore how data engineering drives sustainable business outcomes, while Ghosh and Sarker (2019) propose green IT solutions for sustainable data management. Gupta and Raj (2020) explore machine learning's role in managing energy-efficient data systems, a key facet of green data engineering. Henderson and Lee (2021) emphasize innovative practices in sustainable data storage solutions, and Hussain and Khan (2020) demonstrate how AI-powered data center management can enhance energy efficiency. Jain and Verma (2022) discuss data lifecycle management, emphasizing its importance in optimizing storage and minimizing energy use. Kumar and Sharma (2018) address the intersection of big data and sustainability, presenting key considerations for future technologies. Lewis and Patel (2021) propose strategies for optimizing power usage in cloud environments, contributing to greener data systems. Li and Wang (2020) examine the future of sustainable big data analytics, focusing on the importance of scalable solutions in enterprises. Singh and Agarwal (2021) advocate for sustainable practices in data engineering to reduce environmental impacts, while Zhang and Li (2020) emphasize AI's potential to enhance energy efficiency in data systems. Lastly, Zhao and Zhang (2019) discuss the innovations and challenges in achieving sustainability within the rapidly evolving field of data engineering. Collectively, these studies provide a comprehensive view of the growing importance of sustainability in data engineering, covering various aspects from cloud computing and storage solutions to AI applications and energy-efficient algorithms.

Impact of Data Engineering on the Environment

Data engineering processes, including data storage, processing, and transfer, have a direct and measurable impact on the environment. Key areas of concern include:

Energy Consumption: High-performance servers and data centers require continuous power for operation and cooling.

E-Waste Generation: The frequent upgrade cycles of hardware contribute to electronic waste, much of which is improperly disposed of.

Carbon Emissions: The energy-intensive nature of data operations often relies on non-renewable energy sources, increasing carbon footprints.

Understanding these challenges underscores the urgency of integrating sustainable practices into data engineering workflows.