Chapter 7
Autonomous Data Orchestration With Generative AI:
Redefining Pipelines for Intelligent Analytics

Muhammad Usman Tariq
https://orcid.org/0000-0002-7605-3040
Abu Dhabi University, UAE & University College Cork, Ireland

ABSTRACT

Orchestration of autonomous data using generated AI transforms traditional analytics pipelines by introducing intelligent automation, context-related decision-making, and adaptive data workflows. This ambitious paradigm leverages the capabilities of large-scale models and basic AI systems to dynamically manage the absorption, transformation, integration, and delivery of data, eliminating the need for constant human supervision. Generated AI enables the system to interpret metadata, understand the data's intent, and optimize the pipeline itself based on power metrics or actual time analysis requirements. In contrast to the traditional static architecture, autonomous orchestration introduces continuous learning, enabling the pipeline to evolve further and adapt to changing business requirements and data ecosystems. Enhance agility, eliminate operational bottlenecks, and enhance accessibility for advanced analytics. This shift redefines the role of data engineers, focusing on governance, monitoring, and strategic design while minimizing manual intervention in pipeline operations.

DOI: 10.4018/979-8-3373-5616-7.ch007

INTRODUCTION

The development of data pipelines over the past 20 years has been characterized by a significant transformation of manual code capabilities into intelligent and autonomous systems. In the early stages of data technology, pipelines were created manually using scripting languages and batch-oriented ETL processes (extract, transformation, load). These pipelines were stiff, fragile, and required deep technical knowledge. As data volume, speed, and diversity increased, the limitations of manual engineering became clearer, and orchestration frameworks such as Apache Airflow and Luigi were developed, improving dependency management, planning, and monitoring. However, even these systems remained largely dependent and relied on human intervention (Kurnia et al., 2025). The rise of cloud computing ecosystems, real-time analytics, and API control has given us more complex and necessary pipelines, which are more adaptive, scalable, and resilient. This demand for flexibility has laid the foundation for integrating artificial intelligence into data orchestration. Generation AI introduces new layers of cognitive skills in the orchestration stack, enabling the system to interpret data, generate transformation logic, automate the creation of metadata, and even interact with users in natural language.

Pipelines are no longer limited to specified logic. Now, you can develop dynamically in response to changes in data schemes, business goals, or system anomalies. It represents a fundamental change in the creation and behavior of a data system. Generated AI can analyze the structure and content of data, identify relevant processing steps, generate code, and explain decisions in a manner that aligns with the business context. It fills the gap between data engineering and domain expertise, enabling non-technical users to launch and manage complex workflows with simple input requests (Khokhlov, 2025). This democratization of orchestration encourages greater cooperation, accelerates time, and makes analysis more responsive, adapting to the organization's needs. It examines how traditional pipeline concepts, such as orchestration, planning, transformation, surveillance, and governance, are reinterpreted through the lens of AI. This chapter provides an overview of the fundamental components of intelligent pipeline design, including recording, transformation, observability, and feedback loops. Secondly, the unique functions that the generative model brings to these components include Dynamic schema inference, metadata enrichment, and prompt-based pipeline production. In subsequent sections, we explore architectural patterns, critical tools and platforms, evaluation metrics for AI-controlled orchestration, and key aspects of security, privacy, and governance. This chapter also explores emerging trends, including the self-parenting pipeline and agent-based engineering, offering insights into the ethical and operational implications of fully autonomous systems. Ultimately, this chapter aims to provide readers with a fundamental redesign in a new era of intelligent analytics infrastruc-

Related Content