

# Human-AI Collaboration in Data Migration

A Phase-Wise Reflection on Decision-Making, Automation, and Boundaries

Harishraj Shetty  
Salesforce Consultant,  
Richmond, Virginia, USA

**Abstract**—In large-scale data migration projects, the role of human decision-making is critical—especially when transitioning from legacy systems to modern cloud-based platforms such as Salesforce. This paper consolidates practitioner reflections and observations into a phase-wise review of challenges, decision points, and potential applications of artificial intelligence (AI). It draws on detailed reflections about challenges in data migration, the opportunities AI integration offers, and the lifecycle phases of migration projects. Each phase is described in terms of real-world difficulties, possible AI applications, and boundaries where human oversight is necessary. By traversing the full data migration lifecycle, this paper provides a structured reflection on where AI can enhance outcomes and where human judgment remains irreplaceable.

**Keywords**—Data Migration; Artificial Intelligence (AI); Generative AI (GenAI); Digital Transformation; Project Management; Data Management; Automation; Data Quality; Legacy Systems; Machine Learning (ML);

## I. INTRODUCTION

Data migration is often described as a high-risk, low-visibility activity that organizations tend to postpone because it does not directly create immediate business value. Yet it is one of the most critical enablers for digital transformation, modernization of enterprise systems, and adoption of cloud-based platforms. Migration projects frequently struggle due to hidden complexities, incomplete knowledge of legacy systems, and misaligned expectations among stakeholders. This paper presents a consolidated reflection from practitioner experience to highlight challenges, AI opportunities, and boundaries of human oversight. By organizing observations into a lifecycle perspective, the paper attempts to provide a structured way to think about the role of AI alongside human leadership in migration projects [8][6].

## II. LITERATURE REVIEW

Data platform migrations are inherently challenging, complex, and time-consuming, often plagued by legacy code, data quality issues, and manual effort. Artificial Intelligence (AI) and Generative AI (GenAI) offer transformative solutions, promising to automate and accelerate migrations. Key benefits include [3][17][2]:

Enhanced data quality through profiling, anomaly detection, and pattern recognition [4].

Transparency for legacy code, with GenAI explaining both the mechanics and business context of old scripts, which alleviates the "fear of the unknown" [11].

This enables a "move and improve" approach instead of a simple "lift and shift" [14].

Automated tasks such as data cleansing, mapping, validation, and schema conversion [10].

Machine Learning (ML) models (e.g., XGBoost, Decision Tree) for predictive performance modeling to estimate transfer times and resource needs [13].

Platform migrations are an ideal starting point for AI adoption due to their internal nature, which mitigates public-facing risks and provides clear validation criteria against existing "ground truth" data. This allows organizations to build AI skills in a safe environment, accelerating strategic projects and retaining talent [9].

## III. CHALLENGES IN DATA MIGRATION

The most significant challenge faced by data migration leads is the 'fear of the unknown'. This includes hidden complexity, undocumented knowledge, and unexpected discoveries that surface during later phases. Several challenges are noted across projects [6]:

Lack of subject matter expertise or loss of institutional knowledge when experts leave. Assumptions made early due to missing data that later prove incorrect and derail timelines. Political challenges arise when Subject Matter Experts (SME) raise red flags but lack the influence to be heard. Data complexity, including interconnections, dependencies, and sheer size. Differences in how various stakeholder groups interpret the same data. Difficulty in estimating effort when information is incomplete. The time-consuming and repetitive nature of migration, often sidelined as 'hygiene work'. Project closure difficulties, with complexities unearthed late, and validation becoming a bottleneck. These challenges exist on both the source (legacy) and target (cloud) sides, making data migration doubly complex.

## IV. AI INTEGRATION OPPORTUNITIES

Reflections suggest several points where AI could reduce manual work and support migration teams. These include [4][17]:

AI for automated schema scanning and data profiling to detect data size, type, class, and sensitivity. AI for outlier detection to simplify migration planning and reduce hidden complexity. AI to rapidly understand business domain context, reducing dependency on extensive Subject Matter Expert(SME) sessions. AI-driven enterprise analysis helps with the rapid acquisition of Domain knowledge. Use of AI to create preliminary mappings for simpler data sets, providing a starting point for migration teams. AI-driven enterprise analysis to classify structured and unstructured data and separate simple versus complex schemas. Machine learning applied to resource allocation and optimization of migration configurations. AI-supported reporting and progress tracking during mapping and transformation activities. These reflections emphasize that AI can play a supportive role by accelerating repetitive, predictable, and time-consuming tasks.

## V. DATA MIGRATION LIFECYCLE PHASES

Based on practitioner experience, the following lifecycle phases can be identified:

- A. **Pre-Extraction and Discovery:** This phase involves identifying all unknown complexities, risks, and scope boundaries. It is often characterized by uncertainty, missing knowledge, and reliance on assumptions. AI can help by scanning entire schemas, classifying data, and performing automated profiling to reduce ambiguity [4].
- B. **Planning:** In the planning phase, goals, KPIs, and scope must be clearly defined. Challenges include inaccurate effort estimation, unclear milestones, and misaligned expectations. ML can assist by simulating scenarios, identifying risks, and analyzing dependencies. However, decisions about project strategy, objectives, and KPIs must remain under human control [13].
- C. **Environment Setup:** This involves staging, consolidation, and understanding constraints on source, staging, and target systems. Decisions include whether to bring in master data first or transactional data, and how to manage space and environment limitations. AI can optimize environment planning, but the sequencing of master versus transactional data remains a human responsibility. Further research can be done on how to leverage AI for migration sequencing.
- D. **Data Profiling and Quality:** Profiling data helps understand its characteristics, constraints, and quality levels. AI can accelerate profiling, detect anomalies, and identify outliers. However, human oversight is essential to interpret anomalies in a business context and ensure accuracy.
- E. **Enterprise Analysis and Data Mapping:** This phase involves analyzing functional requirements and aligning them with data mapping needs. AI can provide crude data mappings, anomaly detection, and suggestions based on historical migrations. Yet, final validation of mappings must be done by SMEs, as context and interpretation vary across stakeholders. Leveraging AI tools like NotebookLM can enable enterprise analysis through document analysis for rapid knowledge acquisition.
- F. **Data Transformation, Extract Transform and Load(ETL):** ETL is one of the most resource-intensive phases and is often underestimated in planning. AI could assist by generating transformation rules for simpler datasets and optimizing ETL pipelines. But humans must review all transformations, particularly for sensitive or regulated data [3].
- G. **Data Load and Validation:** At this stage, hidden complexities often surface, making closure difficult. AI can help automate validation, reconciliation, and regression testing. Still, final approval of data validity must come from human leads and business stakeholders [10].
- H. **User Acceptance and Deployment:** Challenges include user resistance and mismatched expectations. AI can generate automated reports and identify high-risk

modules. Nonetheless, acceptance decisions are human-driven, as they require judgment informed by lived experience.

## VI. CONCLUSION

Data migration projects continue to pose significant challenges due to unknown complexities, stakeholder misalignments, and underestimated effort. AI offers promising tools to accelerate discovery, profiling, mapping, and validation, which reduces the manual burden on teams. Reflections drawn from practitioner experience emphasize that while AI can automate repetitive tasks, decisions involving business context, political considerations, and validation must remain with human leads. The path forward is not replacement, but collaboration: AI as a supportive assistant and humans as accountable leaders. Future research could explore the ethical implications of AI in data transformation. More formal case studies are needed to quantitatively measure the Return on investment(ROI) of AI tools in different migration phases. Further research could also explore leveraging AI in data cleaning and maintenance, which should greatly reduce migration risks. Validating the hypothesis that increased AI usage in data maintenance reduces the need for and risks associated with migration could be a case for future research.

## REFERENCES

- [1] "Accelerating Your Database Migrations Using Gen/AI - Lets Talk About Data," AWS Events, YouTube video.
- [2] "AI Unleashed: Revolutionizing Data Platform Migrations," MBN Solutions, YouTube video.
- [3] "Automated Data Migration: The Future of Data Transfer," Functionize.
- [4] PricewaterhouseCoopers, "AI in Data Migration," PwC.
- [5] PricewaterhouseCoopers, "Automation in Data Migration," PwC.
- [6] PricewaterhouseCoopers, "Overcoming Data Migration Challenges," PwC.
- [7] "Automated Data Migration: The Future of Data Transfer," Functionize.
- [8] "Data Migration: How to Do It like a pro," MOSTLY AI, Sep. 20, 2023.
- [9] "Empowering Data Teams with Agentic AI," Google Cloud Events, YouTube video.
- [10] "Enhancing Migration Procedures with AI and Agent-Supported Pre and Post Checks," Cisco Automation Developer Hub, YouTube video.
- [11] DryvIQ, "Intelligent Data Migration: What It Is and Why You Need It," DryvIQ, YouTube video, Jul. 1, 2024.
- [12] "GEN AI Webinar - Data Migration, Development and Integration - Part 1," SID Global Solutions, YouTube video.
- [13] H. Ghaneshirazi, F. Hamouda, M. Fokaefs, W. Haouari, and D. Jania, "DMML: A Machine-Learning Performance Model for Data Migration," in Companion of the 16th ACM/SPEC International Conference on Performance Engineering, New York, NY, USA, 2025, pp. 136–143. doi: 10.1145/3680256.3721313.
- [14] "Navigating Data Migration in the Age of AI: Strategies & Challenges | Nagels Consulting," Nagels Consulting, YouTube video.
- [15] yorku-ease, Yorku-Ease/DataMigrationBenchmarkingTool, GitHub, 2025.
- [16] Hopp Tech, "Navigating Data Migration in the Age of Artificial Intelligence | Hopp Tech Blog," Hopp Tech.
- [17] "Simplify Data Migration with AI-Driven Frameworks | Datafold," Datafold.