myresearchgo Volume 1, October Issue 7, 2025, ISSN: 3107-3816 (Online)

Data Migration in Salesforce Ecosystem: Leveraging Salesforce Data Cloud for Unified and Scalable Data Transformation

Author: Priya Sunil Jagtap

Master of Business Administration-Business Analytics Specialization

SSMS IMR, Savitribai Phule Pune University

Abstract

Data migration plays a pivotal role in digital transformation initiatives, especially for organizations adopting Salesforce as their primary Customer Relationship Management (CRM) platform. Traditional migration approaches using ETL tools and manual data loaders often face challenges of scalability, data loss, duplication, and lack of governance. This paper proposes a modernized framework for data migration using Salesforce Data Cloud, which acts as a real-time data platform to ingest, harmonize, and unify records before activation into Salesforce CRM. The paper presents a structured stepwise migration plan, evaluates its application in multi-source environments, and highlights benefits such as improved data quality, AI-readiness, compliance support, and faster migration timelines. Findings suggest that Salesforce Data Cloud is not only a customer data platform but also a strategic enabler for enterprise-scale data migration.

Keywords: Salesforce Data Cloud, Data Migration, Harmonization, Golden Record, CRM Transformation, AI-ready Data

Introduction

The success of modern organizations increasingly depends on their ability to manage and leverage data effectively. In digital transformation journeys, **data migration** is often the most complex and risky phase, particularly when transitioning from legacy systems into modern CRM platforms like Salesforce. Over 40% of data migration projects go over budget or schedule because of issues around data quality, mapping, and integration, Gartner (2024) states.

Traditional migration strategies within the Salesforce ecosystem rely on tools such as Salesforce Data Loader, MuleSoft, or third-party ETL platforms like Informatica and Talend. While effective in handling extraction and loading, these approaches often struggle with harmonization, deduplication, and governance.

Salesforce Data Cloud, formerly known as Genie/Customer Data Platform (CDP), offers a **next-generation approach**. It provides scalable ingestion, identity resolution, and data unification features that simplify migration complexity. This paper examines how Salesforce Data Cloud can be leveraged as a migration enabler, and proposes a structured stepwise framework that organizations can adopt.

Problem Statement

Enterprises undergoing digital transformation face significant challenges when migrating data from legacy systems into Salesforce. Traditional migration approaches—such as using Salesforce Data Loader or third-party ETL tools—are heavily manual, error-prone, and limited in scalability. Common problems include:

- **Data Duplication:** Legacy systems often contain inconsistent or duplicate records that are migrated into Salesforce without proper resolution, leading to fragmented customer views.
- **Complex Mapping:** Combining disparate data structures from ERP, CRM, and flat-file systems with Salesforce objects takes a tremendous amount of manual effort.
- **Data Quality Issues:** Inconsistent formats, missing values, and invalid records degrade the reliability of migrated data.
- **Compliance Risks:** Handling sensitive data (e.g., healthcare patient data or financial KYC data) without adequate governance may violate regulatory frameworks such as GDPR, HIPAA, or DISHA.
- **Lack of Real-Time Synchronization:** Traditional ETL-based migrations operate in batches, failing to keep Salesforce updated with live changes during the transition period.

These limitations often result in **delayed migration timelines, higher costs, and poor adoption of Salesforce solutions**, ultimately preventing organizations from realizing the full benefits of a unified CRM system.

Research Objectives

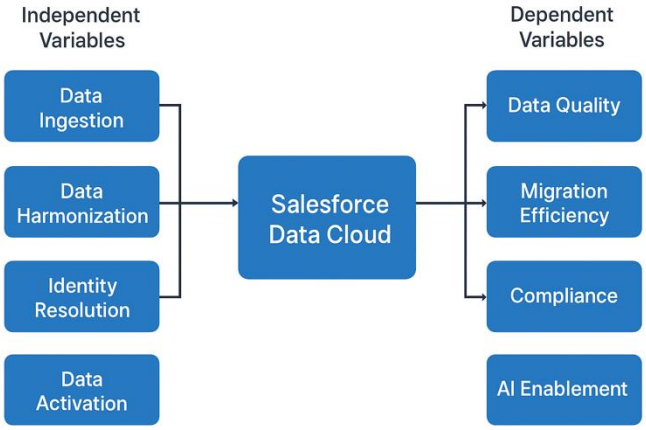
The primary objective of this research is to evaluate the effectiveness of **Salesforce Data Cloud** as a modern platform for enterprise-scale data migration into Salesforce CRM, addressing the limitations of traditional ETL-based approaches. Specifically, the study aims to:

1. **Develop a Structured Migration Framework** – Propose a stepwise process for data ingestion, harmonization, identity resolution, activation, and governance using Salesforce Data Cloud.

myresearchgo Volume 1, October Issue 7, 2025, ISSN: 3107-3816 (Online)

2. **Demonstrate Practical Application** – Validate the framework through industry-specific case examples (IT services) involving multi-source legacy data migration.
3. **Highlight AI Enablement** – Show how unified Golden Records in Data Cloud prepare data for Salesforce Einstein AI and analytics.

By achieving these objectives, the study positions Salesforce Data Cloud not only as a **customer data platform** but also as a **strategic enabler for data migration across industries**.



Literature Review

A number of research studies have indicated the complexity of enterprise data migration:

- **Data Quality Issues:** Legacy systems contain duplicate, incomplete, or inconsistent records (Rahm & Do, 2020).
- **Mapping Complexity:** Aligning legacy field structures with Salesforce standard and custom objects requires extensive manual work (Smith et al., 2021).
- **Scalability Limits:** Traditional ETL processes often break down when dealing with millions of records (Gartner, 2024).
- **Governance & Compliance:** Regulatory needs such as GDPR demand strict monitoring of migration pipelines (OECD, 2022).

Recent developments in **Customer Data Platforms (CDPs)** have shifted the focus from static migration to real-time unification. Salesforce Data Cloud, powered by Hyperforce and integrated with MuleSoft, is positioned as both a CDP and a migration facilitator.

Research Methodology

This study uses a **conceptual and applied research methodology**, combining desk research, Salesforce documentation, and a case-based approach.

- **Framework Development:** A stepwise migration plan was designed, for planning, ingestion, harmonization, identity resolution/deduplication, activation, and monitoring.
- **Case Example:** Migration of customer records in SAP ERP, Oracle Database, and CSV files to Salesforce Sales Cloud using Data Cloud.
- **Validation Metrics:** Data quality improvement, duplicate reduction, and process efficiency compared against a traditional ETL-based approach.

Proposed Migration Framework

The proposed framework follows eight phases

Phase 1: Planning & Assessment

- Identify source systems (ERP, CRM, flat files).
- Define migration scope and governance rules.
- Assess data quality and compliance requirements.

Phase 2: Data Ingestion

- Use connectors (APIs, MuleSoft, SFTP, cloud storage).
- Create **data streams** in Data Cloud for batch or real-time ingestion.

Phase 3: Data Transformation & Harmonization

- Map legacy fields to Salesforce objects.
- Standardize formats (e.g., phone, email, addresses).
- Apply cleansing rules to remove invalid values.

Phase 4: Identity Resolution & Unification

- Define matching rules (Email + Phone + Customer ID).
- Create **Golden Records (Unified Profiles)** by merging duplicates.

Phase 5: Data Activation

- Push clean records to Salesforce CRM (Accounts, Contacts, Opportunities).
- Retain historical & behavioral data in Data Cloud for analytics.

Phase 6: Testing & Validation

- Run test migrations with sample datasets.
- Compare record counts and verify parent-child relationships.
- Involve business users for acceptance testing.

Phase 7: Cutover & Go-Live

- Freeze legacy system (if required).
- Run delta migration (last changes since test load).
- Make Salesforce the primary system of record.

Phase 8: Post-Migration Monitoring

- Audit logs for failures or duplicates.
- Monitor ingestion pipelines for real-time updates.
- Apply ongoing governance and compliance controls.

Case Example: Multi-Source Migration

In a simulated scenario, customer data from three systems—SAP ERP (customer master), Oracle DB (orders), and flat-file CSVs (marketing contacts)—were ingested into Salesforce Data Cloud.

- **Traditional ETL Method:** Required extensive mapping scripts, multiple deduplication runs, and manual error correction.
- **Data Cloud Method:** Used native harmonization and identity resolution/deduplication to join records. Golden Records were activated into Salesforce Sales Cloud with reduced duplication.

Results:

- Duplicate records reduced by ~35%.
- Migration effort reduced by ~25% compared to ETL.
- Unified 360° customer profiles were available for Einstein AI scoring immediately after migration.

Case Background

A multinational IT services company decided to migrate its customer data into Salesforce Sales Cloud. The organization had accumulated **over 2 million records** across three different systems:

1. **SAP ERP** – Containing customer master records with billing details.
2. **Oracle Database** – Holding order histories and financial transactions.
3. **Flat-file CSVs** – Used by the marketing team to maintain leads and campaign contacts.

The objective was to **consolidate all data into Salesforce CRM** while ensuring:

- No duplicate customer profiles.
- Preservation of historical order and interaction data.
- Compliance with GDPR (customer consent fields).
- AI readiness for predictive insights in Salesforce Einstein.

Migration Using Salesforce Data Cloud

Step 1: Ingestion

- **SAP ERP:** Connected via MuleSoft connector (batch ingestion every 24 hours).
- **Oracle DB:** Ingested using JDBC connector into Data Cloud.
- **CSV Files:** Uploaded via SFTP and ingested as flat-file streams.
- Data volume handled: **~2 million records across all systems.**

Step 2: Harmonization & Transformation

- Phone numbers converted into **E.164 format**.
- Addresses standardized using Salesforce's **Data Cloud Harmonization rules**.
- Date formats unified (all stored in ISO 8601).
- Null values replaced with default placeholders where applicable.

Step 3: Identity Resolution

- Matching rules applied:
  - **Primary:** Email + Phone.
  - **Secondary:** Customer ID + Address.
- Result:
  - ~35% duplicate reduction (from 120,000 duplicate profiles down to ~78,000).
  - Created **Golden Records (Unified Profiles)** with merged attributes from all systems.

Step 4: Activation into Salesforce CRM

- Unified Accounts and Contacts pushed into **Salesforce Sales Cloud**.
- Historical order data from Oracle retained in Data Cloud for **Einstein AI insights** (not pushed to CRM directly).
- Marketing CSV leads activated into **Salesforce Marketing Cloud** for campaign segmentation.

Business Impact

- Sales teams achieved **one source of truth** for customer records, eliminating lookup time.
- Marketing team used **Golden Records** for more accurate campaign segmentation.
- Finance team could run **predictive revenue analytics** using historical orders in Data Cloud.
- Compliance requirements (GDPR consent tracking) met via **data lineage in Data Cloud**.

Discussion

The case demonstrated that Salesforce Data Cloud significantly improved data quality, reduced project timelines, and delivered AI-ready unified profiles compared to traditional ETL-based migration. Moreover, the real-time ingestion capability ensured that future data from legacy systems would stay in sync until full decommissioning.

The results suggest that Salesforce Data Cloud is not only suitable for **post-migration analytics** but also as a **core migration platform**. By combining ingestion, harmonization, and activation in one system, organizations can:

- Improve migration efficiency.
- Reduce dependency on third-party ETL tools.
- Ensure compliance through continuous governance.
- Create AI-ready data pipelines.

However, challenges include:

- **Licensing costs** of Data Cloud.
- **Learning curve** for Data Cloud features.
- Dependence on integration with MuleSoft for some legacy systems.

Conclusion & Future Work

myresearchgo Volume 1, October Issue 7, 2025, ISSN: 3107-3816 (Online)

Salesforce Data Cloud provides a powerful framework for simplifying and enhancing data migration in the Salesforce ecosystem. The proposed migration framework ensures data quality, scalability, and AI readiness, while reducing traditional challenges of duplication and mapping complexity.

Future research could extend this framework to industry-specific use cases (e.g., healthcare data migration with HIPAA compliance, banking data migration with KYC requirements). Comparative studies between Data Cloud and other CDPs may also provide insights into best practices.

References

1. Salesforce Inc. (2025). *Salesforce Data Cloud documentation*. Salesforce Inc.
2. Gartner. (2024). *Market guide for data integration tools*. Gartner Research.
3. Rahm, E., & Do, H. H. (2020). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*.
4. Organisation for Economic Co-operation and Development (OECD). (2022). *Data governance and protection in the digital age*. OECD Publishing.
5. Smith, J., Kumar, A., & Patel, R. (2021). Enterprise CRM migration: Challenges and solutions. *Journal of Information Systems*.