

DATA MIGRATION STRATEGY

Core Digital Transformation for Retail Banking

Legacy System Migration to Cloud-Native Microservices Architecture

Project	Prepared By	Version	Date
Core Digital Transformation	Data Engineering Lead / Enterprise Architect	1.0 — Initial Release	February 18, 2026

CONFIDENTIAL — FOR INTERNAL USE ONLY

1. Executive Summary

This Data Migration Strategy defines the end-to-end approach for migrating all critical banking data from legacy siloed systems to the new cloud-native, microservices-based Enterprise Data Warehouse (EDW) and operational databases. The migration covers four source systems — CRM, Loan Management, Core Banking, and Transaction Logs — and must be executed with zero data loss, full regulatory compliance (KYC, AML, PCI-DSS, GDPR), and minimal business disruption.

The strategy adopts a phased, parallel-run migration model with automated validation checkpoints, rollback capabilities at every phase, and full audit traceability from source to target. A Conditional Go/No-Go decision gate is embedded at each phase boundary before production cutover is authorized.

2. Migration Objectives

ID	Objective	Domain	Priority
OBJ-01	Migrate 100% of customer, account, loan, and transaction records from legacy systems to the new EDW with zero data loss.	Data Integrity	CRITICAL
OBJ-02	Maintain full audit traceability — every record traceable from source system through staging to the target EDW.	Compliance	CRITICAL
OBJ-03	Complete migration within the approved 12-week timeline with production cutover no later than Q3 2026.	Schedule	HIGH
OBJ-04	Ensure all migrated data satisfies KYC, AML, PCI-DSS, and GDPR regulatory controls before go-live sign-off.	Regulatory	CRITICAL
OBJ-05	Minimize production system downtime to approved maintenance windows only; target zero unplanned downtime.	Operations	HIGH
OBJ-06	Standardize data formats, customer IDs, and currency codes across all migrated datasets.	Data Quality	HIGH

OBJ-07	Implement automated rollback procedures at each migration phase, executable within 30 minutes of failure detection.	Risk Management	HIGH
--------	---	-----------------	------

3. Migration Scope

3.1 In Scope

Source System	Data Domain	Estimated Volume
CRM System	Customer profiles, KYC records, contact history, marketing preferences, and service interactions	~2.4M customer records
Loan Management	Loan applications, repayment schedules, delinquency records, interest rate history, and credit scoring data	~850K loan records
Core Banking System	Account balances, account lifecycle events, customer IDs, overdraft records, and product configurations	~3.1M account records
Transaction Logs	Card payments, ATM withdrawals, online transfers, direct debits, and historical fraud alert records	~180M transaction rows
Compliance Archives	AML audit trails, KYC event logs, GDPR data subject requests, and regulatory reporting history	~12M audit log entries

3.2 Out of Scope

- Mobile application data — addressed in a separate mobile workstream
- Third-party vendor proprietary data not owned by the bank
- International subsidiary systems — covered under a separate migration plan
- Physical branch infrastructure and ATM firmware data
- Email and document management systems (SharePoint, DMS)

4. Source System Assessment

A comprehensive assessment of each legacy source system was conducted to identify data quality issues, structural inconsistencies, and compliance gaps prior to migration planning. Key findings are summarised below.

Source System	Technology	Key Issues Identified	Data Quality Risk	Remediation Required
CRM	Oracle DB 11g	Duplicate customer IDs; inconsistent phone number formats; 8% null KYC fields	HIGH	Deduplication; KYC field enrichment; ID standardisation
Loan Management	IBM DB2 v10	Interest rate precision mismatch; loan_type enum inconsistency; 2.3% orphaned loan records	MEDIUM	Enum normalisation; orphan record resolution; precision conversion
Core Banking	Temenos T24	Multi-currency accounts with incorrect base currency flags; closed accounts still marked active	HIGH	Currency code standardisation; account status reconciliation
Transaction Logs	MS SQL Server 2014	Missing reference_no on 0.4% of records; timestamp timezone inconsistency (UTC vs local)	MEDIUM	Reference number generation; timezone normalisation to UTC
Compliance Archives	Flat files / CSV	Unstructured format; no primary keys; AML flags not machine-readable	CRITICAL	Schema imposition; primary key generation; AML flag parsing

5. Migration Approach & Strategy

5.1 Overall Strategy: Phased Parallel-Run Migration

The migration follows a phased parallel-run model. Legacy source systems remain fully operational throughout all migration phases. Migrated data in the new EDW is validated in parallel against live source system data before any cutover is authorized. This eliminates risk of data loss and provides a clear rollback path at every stage.

Migration Flow:

Legacy Source Systems → ETL Extract → Staging Area → Data Validation → Transform & Cleanse → Load to EDW → Parallel Validation → Go/No-Go Gate → Production Cutover

Rollback available at every arrow. Legacy systems remain live until final cutover is signed off.

5.2 ETL Pipeline Architecture

- **Extract: Incremental extraction via Change Data Capture (CDC) for real-time systems; full-load extraction for compliance archives**
 - CRM and Core Banking: CDC using Debezium connector over JDBC
 - Transaction Logs: Bulk export via SQL Server BCP then CDC for deltas
 - Compliance Archives: Batch ingestion via Apache NiFi with schema mapping
- **Transform: Apache Spark jobs executed on IBM Cloud Pak for Data**
 - Customer ID standardisation across all source systems
 - Date/time normalisation to ISO 8601 UTC format
 - Currency code standardisation to ISO 4217
 - PII masking for non-production environments (GDPR compliance)
 - Duplicate detection and resolution using deterministic matching rules
- **Load: Incremental upsert to PostgreSQL EDW using Apache Kafka + Kafka Connect sink**
 - Upsert logic: INSERT on new records, UPDATE on changed records (keyed on UUID)
 - Data partitioned by date for query performance and archival management
 - Post-load record count and hash validation before phase sign-off

5.3 Data Extraction Methods by Source

Source	Extraction Method	Tool	Frequency	Encryption
CRM	CDC via Debezium JDBC	Debezium + Kafka	Real-time stream	TLS 1.3 in transit
Loan Management	Full load + CDC delta	Apache NiFi	Daily batch + live CDC	TLS 1.3 in transit
Core Banking	CDC via Debezium JDBC	Debezium + Kafka	Real-time stream	TLS 1.3 in transit
Transaction Logs	BCP bulk export + CDC	SQL BCP + Kafka Connect	Hourly batch + live CDC	TLS 1.3 in transit
Compliance Archives	Batch CSV ingestion	Apache NiFi + Spark	One-time full load	AES-256 at rest

6. Migration Phases & Timeline

Phase	Name	Activities	Duration	Owner
Phase 0	Preparation & Environment Setup	Provision cloud infrastructure; configure ETL tools; establish connectivity to source systems; baseline data profiling	Weeks 1–2	Enterprise Architect
Phase 1	Compliance Archives Migration	Ingest and parse flat-file compliance archives; impose schema; generate PKs; validate AML flag integrity; load to EDW compliance_logs table	Weeks 3–4	Data Engineering Lead
Phase 2	Customer & Account Migration	Extract CRM and Core Banking data; deduplicate customers; standardise IDs; run KYC validation; load customers and accounts tables; parallel run validation	Weeks 5–7	Data Engineering Lead
Phase 3	Loan Portfolio Migration	Extract Loan Management data; resolve orphan records; normalise loan_type enums; credit score mapping; load loans and credit_scores tables	Weeks 7–8	Data Engineering Lead
Phase 4	Transaction History Migration	Bulk export + CDC for transaction logs; UTC timezone normalisation; reference number generation for gaps; load transactions table; integrity hash check	Weeks 8–10	Data Engineering Lead
Phase 5	Parallel Run & UAT Validation	Run legacy and new systems in parallel; reconcile record counts and checksums; perform UAT with business units; regulatory compliance sign-off from CCO	Weeks 10–11	Business Analyst / CCO
Phase 6	Production Cutover & Decommission	Final Go/No-Go gate; switch traffic to new platform; monitor for 72 hours; legacy systems placed in read-only archive mode; decommission scheduled	Week 12	CIO / Enterprise Architect

7. Data Validation & Quality Framework

7.1 Validation Checkpoints

Automated validation is executed at three points in the ETL pipeline: post-extraction, post-transformation, and post-load. All validation results are logged with timestamps, record counts, and failure details for audit purposes.

Checkpoint	Stage	Validation Rules Applied	Pass Threshold	On Fail
VAL-01	Post-Extraction	Record count vs source; null check on mandatory fields; referential integrity; duplicate primary key detection	100%	HALT
VAL-02	Post-Transformation	Customer ID uniqueness; date format ISO 8601; currency code ISO 4217; KYC status enum valid; PII masking applied in non-prod	99.9%	REVIEW
VAL-03	Post-Load (EDW)	Record count reconciliation source vs EDW; MD5 hash comparison on key fields; foreign key integrity; compliance field completeness	100%	ROLLBACK
VAL-04	Parallel Run	Business logic reconciliation: account balances match, loan status consistent, transaction totals agree within 0.01% tolerance	99.99%	NO-GO

7.2 Data Quality Rules

Rule ID	Rule Name	Description	Applies To
DQ-01	Completeness Check	All mandatory fields (PK, NOT NULL constraints) must have values. Zero null tolerance on critical fields.	All tables
DQ-02	Uniqueness Check	No duplicate primary keys. Customer email must be unique across the customer table.	customers, accounts
DQ-03	Referential Integrity	All foreign keys must resolve to existing parent records. Orphan records flagged and quarantined.	accounts, loans, transactions
DQ-04	Format Standardisation	Dates in ISO 8601; currency in ISO 4217; phone numbers in	All tables

		E.164 format; UUIDs in RFC 4122.	
DQ-05	Business Rule Validation	Loan principal > 0; account balance >= 0 for savings; KYC status in allowed enum values.	loans, accounts
DQ-06	PII Compliance	PII fields (name, DOB, email, phone) masked in non-production environments per GDPR Article 25.	customers
DQ-07	AML Flag Integrity	All compliance log entries with severity=CRITICAL must have a non-null event_type and entity_id.	compliance_logs

8. Security & Regulatory Compliance

8.1 Encryption Standards

Control	Standard	Application
Encryption in Transit	TLS 1.3	All data flows between source systems, ETL pipeline, Kafka, and the target EDW are encrypted with TLS 1.3
Encryption at Rest	AES-256	Staging Area, Enterprise Data Warehouse, and archive storage encrypted at rest using AES-256 with bank-managed keys
Key Management	IBM Secrets Manager / AWS KMS	All encryption keys managed by centralised key management service. No keys hardcoded in ETL scripts or configuration files
Data Masking	Tokenisation + Redaction	PII fields tokenised in non-production environments. Cardholder data isolated in PCI-DSS compliant zone with field-level redaction for non-payment systems

8.2 Regulatory Compliance Mapping

Regulation	Requirement	Migration Control	Owner
KYC	Customer identity verified before account activation	KYC status field migrated with full history; verification timestamps preserved; rejected records quarantined for review	Compliance Officer
AML	Transaction monitoring logs retained and auditable	All compliance_log entries migrated with severity ratings intact; AML flags preserved as machine-readable enum values	Compliance Officer
PCI-DSS	Cardholder data isolated; access restricted	Payment card data migrated to isolated PCI-DSS zone; RBAC enforced; no cardholder data in general-purpose tables	CISO
GDPR	PII handled lawfully; right to erasure supported	PII fields tagged in EDW schema; data subject access requests (DSAR) supported via customer_id lookup; deletion workflows tested pre-cutover	CCO / DPO

8.3 Access Controls During Migration

- All ETL service accounts operate with least-privilege permissions scoped to specific source tables and target schemas
- Migration environment access logged in compliance_logs with user ID, timestamp, and action performed

- Production source systems accessible only via read-only migration service accounts during extraction
- Multi-factor authentication (MFA) enforced for all engineers accessing the migration environment
- Privileged access reviews conducted daily during active migration phases by the CISO

9. Rollback Strategy

A structured rollback plan is maintained at each migration phase. Rollback can be initiated within 30 minutes of failure detection. Legacy source systems are kept in a fully operational state until the final production cutover is signed off by the CIO.

Phase	Rollback Trigger	Rollback Action	Time to Execute	Decision Authority
Phase 0	Infrastructure provisioning failure	Deprovision cloud resources; revert to pre-migration state	< 15 minutes	Enterprise Architect
Phase 1	Compliance archive load failure or AML flag corruption detected	Truncate compliance_logs staging table; re-run full ingestion from source archives	< 30 minutes	Data Engineering Lead
Phases 2–4	Record count mismatch > 0.1%; referential integrity failure; UAT rejection	Restore EDW tables from pre-phase snapshot; notify stakeholders; root cause analysis before re-attempt	< 30 minutes	Data Engineering Lead + CCO
Phase 5	Parallel run reconciliation fails; business unit UAT sign-off withheld	Extend parallel run period; reprocess affected data segments; escalate to CIO	< 2 hours	CIO
Phase 6 (Cutover)	Post-cutover production incident; P1 data integrity failure	Switch DNS/load balancer back to legacy systems; new platform placed in maintenance mode; full incident investigation	< 15 minutes	CIO + CEO

10. Migration Risk Register

ID	Risk Description	Probability	Impact	Mitigation Strategy	Owner
R-01	Data loss during bulk extraction from legacy systems due to network timeout or source system failure	LOW	CRITICAL	Incremental CDC with offset tracking; automated retry with exponential backoff; extraction checkpoints every 100K records	Data Eng Lead
R-02	Referential integrity violations — orphaned records in loan or transaction tables after migration	MEDIUM	HIGH	Pre-migration orphan record analysis; quarantine table for unresolved orphans; manual review workflow before load	Data Eng Lead
R-03	Regulatory non-compliance — KYC or AML data incorrectly transformed or missing after migration	LOW	CRITICAL	Dedicated compliance validation pipeline with CCO sign-off at each phase; automated AML flag integrity checks; audit trail preserved end-to-end	CCO
R-04	Extended migration timeline causing business disruption or missed regulatory deadlines	MEDIUM	HIGH	Phased approach with independent phase delivery; buffer weeks built into schedule; weekly CIO progress reviews	CIO
R-05	Legacy source system performance degradation during parallel CDC extraction impacting live banking operations	MEDIUM	HIGH	CDC extraction throttled to off-peak hours; read-only replica databases used for bulk extraction; source system performance monitored continuously	Enterprise Architect
R-06	Security breach during migration — interception of PII or cardholder data in transit	LOW	CRITICAL	TLS 1.3 end-to-end; PII masked before non-prod exposure; zero-trust network policies; CISO daily access reviews during migration	CISO
R-07	Data format mismatch between legacy enums and new EDW CHECK constraints causing load failures	HIGH	MEDIUM	Pre-migration enum mapping exercise completed for all source systems; transformation rules tested against 10% data sample in sandbox	Data Eng Lead

11. Roles & Responsibilities

Role	Stakeholder	Migration Responsibilities
Migration Owner	CIO	Overall accountability for migration delivery; Go/No-Go authority at production cutover; executive escalation point
Data Engineering Lead	Data Engineering Lead	ETL pipeline design and execution; data validation framework; phase delivery; rollback execution; post-load reporting
Enterprise Architect	Enterprise Architect	Migration architecture decisions; source system connectivity; infrastructure provisioning; integration governance
Compliance Sign-off	CCO	KYC/AML/GDPR validation sign-off at each phase; regulatory reporting readiness confirmation; Go/No-Go co-approval
Security Oversight	CISO	Encryption controls validation; access review during migration; PII masking verification; security incident response
UAT Coordination	Business Analyst	UAT test plan design; business unit coordination; parallel run reconciliation sign-off; defect tracking
Independent Validation	External Auditor	Independent compliance validation at Phase 5; audit evidence review; regulatory submission support

12. Monitoring & Audit Logging

12.1 Migration Monitoring KPIs

KPI	Target	Alert Threshold	Monitoring Tool
Record extraction rate	>= 50,000 records/min	< 30,000 records/min	Kafka Consumer Lag Monitor
ETL pipeline latency	< 500ms per batch	> 1,000ms per batch	IBM Cloud Monitoring
Validation failure rate	< 0.01% of records	> 0.1% of records	Custom validation dashboard
Record count reconciliation	100% match source vs EDW	Any discrepancy	Reconciliation report (auto)
Source system CPU/IO impact	< 20% additional load	> 35% additional load	Datadog / Dynatrace

12.2 Audit Log Requirements

- Every ETL batch execution logged: timestamp, source system, record count extracted, validation result, and operator ID
- Every transformation rule applied logged with rule ID, field affected, before/after values, and batch ID
- Every load operation logged: target table, record count inserted/updated, hash validation result, and load duration
- All rollback events logged with trigger condition, action taken, records affected, and authorisation details
- Audit logs retained for 7 years in the EDW compliance_logs table in accordance with regulatory requirements
- Audit logs are immutable — write-once, append-only with no update or delete permissions on compliance_logs

13. Go / No-Go Decision Gates

A formal Go/No-Go decision gate is conducted at the end of each migration phase before proceeding to the next. All criteria must be met for a GO decision. A NO-GO triggers a defined remediation period before re-assessment.

Gate	Go Criteria	Decision Authority	Max Wait
Gate 1 Phase 1→2	100% compliance archives loaded; AML flags validated; compliance_logs record count matches source archive	CCO + Data Eng Lead	48 hours
Gate 2 Phase 2→3	Customer deduplication complete; KYC status preserved on all records; accounts FK integrity 100%; parallel read validation passes	CCO + CIO	48 hours
Gate 3 Phase 3→4	All loan records loaded; no orphaned loans; credit scores mapped correctly; loan status enums validated; business unit sign-off	Data Eng Lead + BA	48 hours

Gate 4 Phase 4→5	Transaction count reconciles within 0.01% of source; UTC timestamps consistent; reference_no unique constraint passes; no data loss confirmed	CIO + CISO	72 hours
Gate 5 CUTOVER	All parallel run reconciliations pass; UAT signed off by all business units; CCO regulatory compliance confirmed; CISO security controls validated; External Auditor review complete	CEO + CIO + CCO	Exec decision

14. Strategy Validation Checklist

Validation Criteria	Status
All source systems identified with data volumes and quality risk ratings	YES
Migration objectives defined with priority levels	YES
ETL pipeline architecture documented (Extract → Transform → Load)	YES
Data extraction method defined per source system	YES
Phased migration timeline with owners defined (6 phases)	YES
Data validation framework with 4 checkpoints and pass/fail thresholds	YES
Data quality rules documented (DQ-01 through DQ-07)	YES
Encryption standards and key management documented	YES
KYC, AML, PCI-DSS, and GDPR regulatory compliance controls mapped	YES
Rollback strategy defined per phase with time targets	YES
Risk register documented with probability, impact, and mitigations (7 risks)	YES
Roles and responsibilities assigned across all migration functions	YES
Migration KPIs and audit log requirements defined	YES
Go/No-Go decision gates defined with criteria and decision authority	YES

15. Document Approval & Sign-Off

By signing below, the designated approvers confirm they have reviewed this Data Migration Strategy and authorize its implementation as part of the Core Digital Transformation for Retail Banking programme.

Name	Title	Signature	Date
	Chief Information Officer (CIO)		
	Chief Compliance Officer (CCO)		

	Chief Information Security Officer (CISO)		
	Enterprise Architect		
	Data Engineering Lead		