# DATA FLOWCHART

## Consolidation of Data Infrastructure

Core Digital Transformation for Retail Banking

| | |
|---|---|
| **Project** | Core Digital Transformation for Retail Banking |
| **Document Type** | Data Flow Diagram (DFD) – Enterprise Data Consolidation |
| **Date** | February 18, 2026 |
| **Prepared By** | Data Engineering Lead / System Architect |
| **Version** | 1.0 – Initial Release |
| **Purpose** | Visualize end-to-end data integration from source systems to enterprise data warehouse |

## 1. Overview

This data flowchart represents the consolidation of fragmented data systems across the retail bank into a unified, enterprise-grade data infrastructure. The diagram shows how data flows from four critical source systems (CRM, Loan Management, Core Banking, and Transaction Logs) through a secure ETL pipeline into a centralized Enterprise Data Warehouse (EDW). The architecture includes staging layers, data validation checkpoints, audit logging, encryption controls, and compliance annotations to meet KYC, AML, PCI-DSS, and GDPR requirements.

## 2. Key Data Sources

| Source System | Description |
|---|---|
| **CRM** | Customer profiles, contact history, marketing preferences, and service interactions |
| **Loan Management** | Loan applications, repayments, delinquencies, interest rates, and credit scoring data |
| **Core Banking System** | Account balances, customer IDs, account opening/closing events, and overdraft records |
| **Transaction Logs** | Card payments, transfers, ATM withdrawals, online banking transactions, and fraud alerts |

## 3. Data Flow Diagram Symbols

| Symbol Type | Shape | Meaning/Use | Example |
|---|---|---|---|
| **Entity (External System)** | Rectangle | Represents an external data source or destination outside the system boundary | CRM, Loan Management, Core Banking, Transaction Logs |
| **Process (ETL Job/Transform)** | Rounded Rectangle or Circle | Represents data transformation, movement, or processing (ETL steps) | Extract Data, Transform Data, Validate, Load to Staging |

| Symbol Type | Shape | Meaning/Use | Example |
|---|---|---|---|
| **Data Store (Database/Warehouse)** | Two Parallel Lines or Cylinder | Represents storage where data is held temporarily or permanently | Staging Area, Enterprise Data Warehouse, Archive Store |
| **Data Flow (Movement of Data)** | Arrow (Labeled) | Indicates the direction of data flow between entities, processes, or stores | CRM → Extract → Transform → EDW (label: customer data, loan data, etc.) |
| **Validation/Audit Checkpoint** | Diamond | Represents decision or validation steps used in some extended DFDs | Data Validation Complete? Audit Log Created? |

## 4. Color Coding Legend

### Source Systems

- Blue: External data sources (CRM, Loan Management, Core Banking, Transaction Logs)

### ETL Pipeline

- Green: ETL processes (Extract, Transform, Load)

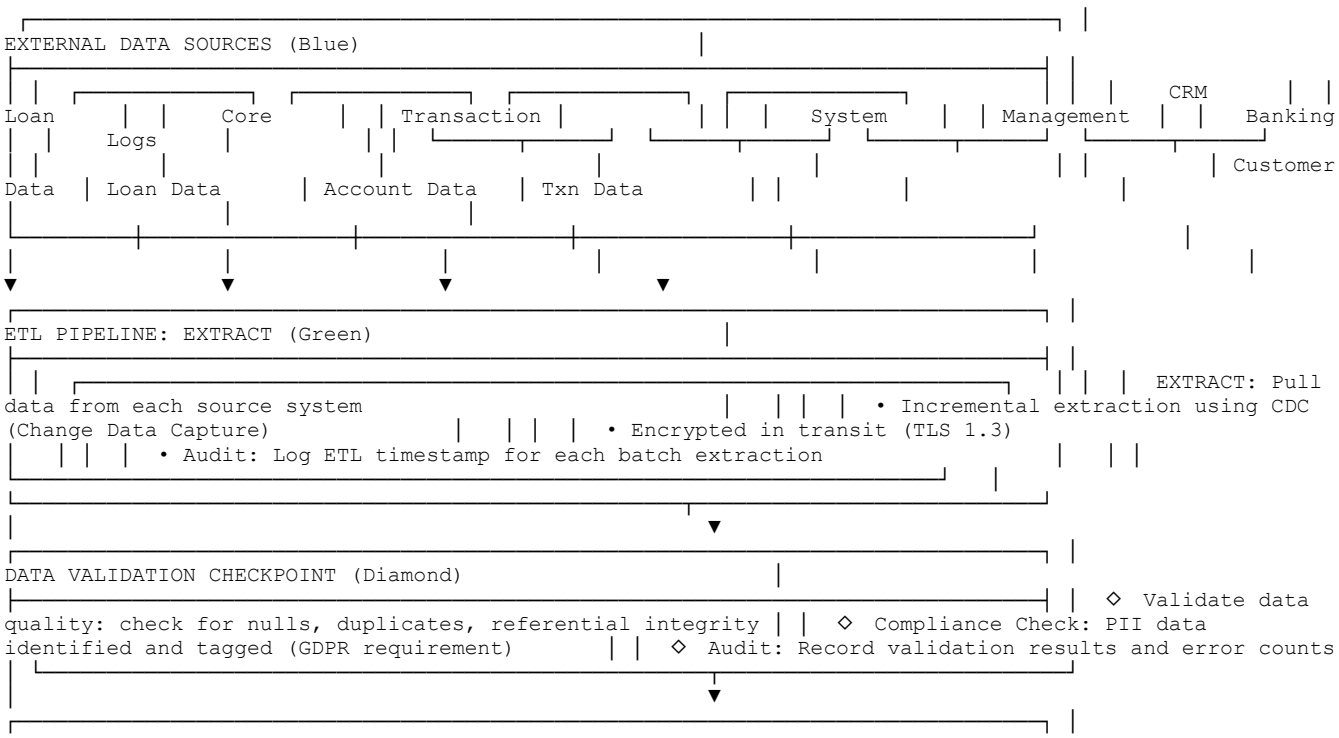
### Storage Layers

- Orange: Data storage layers (Staging Area, Enterprise Data Warehouse, Archive)

### Outputs

- Purple: Final outputs and reporting (Regulatory Reports, Compliance Dashboards, Executive BI/Analytics)

## 5. Data Flowchart Diagram




## 6. Security & Compliance Annotations

### Encryption

- TLS 1.3: All data encrypted in transit between source systems and ETL pipeline
- AES-256: All data encrypted at rest in Staging Area, Enterprise Data Warehouse, and Archive

### Audit Logging

- ETL Timestamp: Every extraction batch logged with timestamp and source system
- Transformation Rules: All data transformations logged with rule ID and applied fields
- Access Logs: All data warehouse queries logged with user ID, timestamp, and data accessed
- Report Generation: All regulatory reports logged with recipient and delivery confirmation

### Regulatory Compliance

- KYC: Customer identity verification data retained for 7 years in EDW

- AML: Transaction monitoring logs retained and flagged transactions auditable
- PCI-DSS: Cardholder data isolated in secure zone with restricted access
- GDPR: PII tagged, masked in non-production, and subject to data subject access/deletion requests

### Data Lineage & Retention

- Data Lineage: Full traceability from source system → staging → EDW → reports
- Staging Retention: 7 days (then purged or moved to archive)
- EDW Retention: 7 years for compliance; older data moved to cold storage or deleted

## 7. ETL Pipeline Details

### Extract Stage

- Source: CRM, Loan Management, Core Banking, Transaction Logs
- Method: Incremental extraction using Change Data Capture (CDC)
- Frequency: Real-time streaming for transactions; daily batch for CRM and loan data
- Security: TLS 1.3 encryption in transit; API authentication with OAuth 2.0

### Transform Stage

- Standardize: Normalize customer IDs, date formats, and currency codes
- Cleanse: Remove duplicates, fix data quality issues, validate referential integrity
- Enrich: Calculate derived fields (e.g., customer lifetime value, risk scores)
- Mask PII: Redact sensitive fields for non-production environments (GDPR)

### Load Stage

- Target: Enterprise Data Warehouse (EDW)
- Method: Incremental load with upsert logic (insert new, update existing)
- Partitioning: Data partitioned by date for query performance
- Validation: Post-load validation checks for record count and data integrity

## 8. Summary

This data flowchart provides a comprehensive view of how fragmented data systems across the retail bank are unified into a single, secure, and compliance-ready enterprise data warehouse. The ETL pipeline includes robust validation checkpoints, encryption controls, and audit logging at every stage to meet KYC, AML, PCI-DSS, and GDPR requirements. The final outputs include regulatory reports, real-time compliance dashboards, and executive analytics systems — all fed from a centralized, trustworthy data layer.