Northstar Analytics

CECL Suite v4.2
Data Requirements, Mapping & ETL Specification

Document ID: NA-CECL-DMES-4.2
Release Date: February 17, 2026

Confidential - For Client Implementation Use Only

Prepared by: Northstar Analytics, Data Engineering & Model Integration
Primary contacts: dataonboarding@northstaranalytics.example |
modelsupport@northstaranalytics.example
Supported delivery methods: SFTP (preferred) or secure REST API (optional)

Document Control

Version	Date	Author	Reviewer	Summary of Changes
1.0	2025-09-05	A. Gomez	L. Nguyen	Initial data onboarding guide and canonical schema definition
1.1	2025-12-02	A. Gomez	L. Nguyen	Expanded recoveries and collateral fields; added QC rule catalog
1.2	2026-02-17	A. Gomez	L. Nguyen	Updated file interface specs; added macro series mapping; clarified reconciliations

Distribution and Use

This specification is intended for implementation teams responsible for extracting, transforming, and delivering source data to the Northstar CECL Suite. It is written in a vendor operational format and does not follow client model documentation templates. The client remains responsible for ensuring that delivered data is complete, accurate, and reconciled to source systems.

Table of Contents (Generated in Word)

This document includes a placeholder for a Microsoft Word-generated Table of Contents. To update: References > Table of Contents > Update Table.

1. Introduction and Integration Overview

Purpose. This specification defines data requirements for the Northstar CECL Suite v4.2, including canonical field definitions, mapping guidance, transformation rules, and quality controls. The objective is to enable repeatable, controlled delivery of monthly (or quarter-end) input data for expected credit loss estimation.

Audience. Primary audiences include data engineering, IT integration, finance controllership support, and model owners responsible for ensuring data lineage and reconciliations.
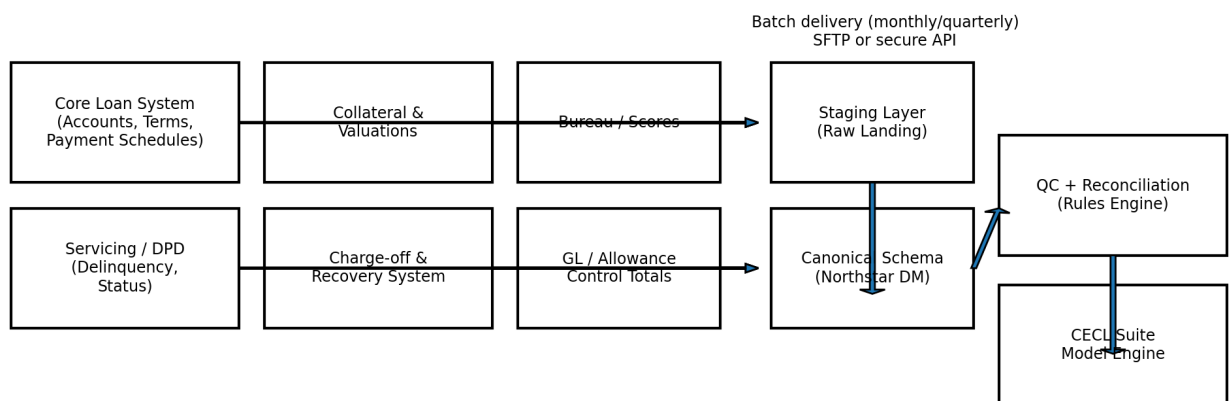
1.1 Integration Options
- Batch SFTP (preferred): client delivers encrypted flat files to a managed landing zone.
- Secure API (optional): client pushes payloads to a REST endpoint with mutual TLS.
- Hybrid: master/static tables via API, snapshots via SFTP.

1.2 Data Delivery Cadence
The standard cadence is a month-end snapshot with an as-of date equal to the reporting cutoff. Transactional feeds (payments, charge-offs, recoveries) may be delivered monthly and are referenced by posting date. Macro scenario series are delivered per scenario set and tied to a named forecast package.

Exhibit 1: High-Level Data Flow

2. Source Systems and Required Extracts

Northstar does not mandate a specific vendor system. The following sources are commonly used. Clients may provide equivalent fields from alternative systems, subject to mapping review.

2.1 Typical Source Systems

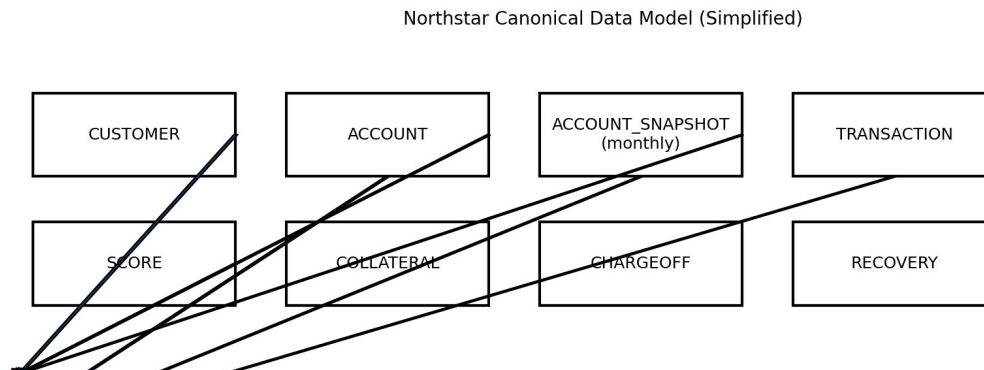| System	Typical Content
Core loan accounting	Account master, terms, balances, rates, schedules
Servicing / collections	Delinquency status, workout flags, repossession status
Charge-off and recovery	Charge-off events, recoveries, collateral liquidation proceeds
Collateral / appraisal	Collateral type, value, valuation dates, lien position
Credit bureau / internal score	Scores and score dates, risk bands, behavioral scores
General ledger	Allowance control totals, charge-off and recovery GL accounts
Reference data	Product codes, branch/channel, portfolio hierarchy

2.2 Required Extracts (Minimum Set)

Extract / Table	Frequency	Grain	Notes
ACCOUNT_MASTER	Monthly	Loan/Line	One row per account at month-end
ACCOUNT_SNAPSHOT	Monthly	Loan/Line	Balances, DPD, status, segment attributes
TRANSACTION	Monthly	Txn	Payments, fees, interest, curtailments
CHARGEOFF	Monthly	Event	Charge-off events and amounts
RECOVERY	Monthly	Event	Recovery cashflows and liquidation proceeds
COLLATERAL	As available	Collateral	Collateral attributes and valuations
SCORE	Monthly/As available	Score	Bureau/internal score observations
GL_CONTROL_TOTALS	Monthly	Portfolio	Allowance and charge-off/recovery totals for reconciliation
MACRO_SERIES	Per scenario set	Time series	Forecast paths for macro variables used in modeling

3. Canonical Data Model (Northstar DM)

The Northstar canonical data model standardizes client data into a consistent schema. This schema is used by the CECL Suite calculation engine and QC rules. Fields are named using snake_case. Dates are ISO-8601 (YYYY-MM-DD). Amounts are in USD unless otherwise specified.

3.1 Entities and Keys

Northstar Canonical Data Model (Simplified)

Primary keys. account_id uniquely identifies an account (loan or credit line). customer_id uniquely identifies a borrower relationship. Snapshots are keyed by (account_id, as_of_date). Transactions are keyed by (account_id, txn_id). All tables include a source_system field for traceability.

3.2 Required Identifiers
- account_id: stable identifier across time; should not be re-used across closed accounts.
- customer_id: stable relationship identifier; may map to CIF or equivalent.
- as_of_date: snapshot date for month-end; must align to reporting cutoff.
- product_code and portfolio_code: reference dimensions used for segmentation and reporting.

4. File Interfaces and Delivery Standards

4.1 File Format

Northstar supports pipe-delimited (*.txt) and CSV (*.csv). UTF-8 encoding is required. Header rows are required and must match field names in this specification. Null values should be represented as empty strings. Date fields must use YYYY-MM-DD. Amount fields must use decimal format with period separator.

4.2 File Naming Convention (SFTP)

Example: <client_code>_<table_name>_<as_of_date>_<load_timestamp>.csv

Example: SRB_ACCOUNT_SNAPSHOT_2026-01-31_20260201T031500Z.csv

4.3 Encryption and Compression

- Files must be encrypted using PGP with a Northstar-provided public key.
- Optional compression: gzip prior to encryption (preferred for large files).
- Checksum: SHA-256 checksum file may be included for integrity verification.

4.4 Delivery Schedule

Standard delivery is within T+3 business days after month-end. If the client performs post-close corrections, Northstar supports re-loads with incremented load_timestamp. Re-loads must include a complete replacement for the affected table and date unless delta mode is explicitly enabled.

5. Field-Level Data Dictionary (Required Tables)

This section defines the minimum required fields for core tables. Clients may provide additional fields; however, additional fields are not consumed unless configured. Req indicates whether the field is required (Y) or optional (N).

5. ACCOUNT_MASTER

Account master attributes and contractual terms (one row per account, static or slowly changing).

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
account_id	string	Y	Unique account identifier	No spaces; <= 64 chars	Stable over time; do not re-use
customer_id	string	Y	Unique borrower relationship identifier	<= 64 chars	Maps to CIF or equivalent
product_code	string	Y	Product code	Reference table	Used for segmentation and reporting
portfolio_code	string	Y	Portfolio hierarchy code	Reference table	E.g., RETAIL_AUTO
origination_date	date	Y	Origination date	YYYY-MM-DD	
maturity_date	date	Y	Contractual maturity date	YYYY-MM-DD	For revolving, contractual term end
interest_rate	decimal	Y	Current interest rate	Decimal (e.g., 0.0725)	Do not provide percent unless configured

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
rate_type	string	N	Fixed/ Variable indicator	FIXED| VARIABLE	
original_bala nce	decimal	N	Original principal balance	>=0	
credit_line	decimal	N	Credit line limit (revolving)	>=0	Required for revolving products
currency	string	N	Currency code	ISO 4217 (e.g., USD)	Default USD if omitted
channel_code	string	N	Origination channel	Reference table	Branch, digital, indirect, etc.
state	string	N	Borrower state	USPS 2-letter	
lien_position	string	N	Lien position (secured)	1|2|3|OTHER	
collateral_id	string	N	Collateral link identifier	<=64 chars	Join to COLLATERAL
status_code	string	Y	Account status	ACTIVE| PAID| CHARGEOFF| CLOSED	As of reporting cutoff

5. ACCOUNT_SNAPSHOT

Month-end snapshot of time-varying account attributes (one row per account per as_of_date).

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
account_id	string	Y	Account identifier		

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
as_of_date	date	Y	Snapshot date	YYYY-MM-DD	Month-end reporting cutoff
current_balance	decimal	Y	Ending balance as of as_of_date	>=0	Principal balance for amortizing; statement balance for revolving
accrued_interest	decimal	N	Accrued interest balance	>=0	If not available, omit
dpd	int	Y	Days past due	>=0	If not available, provide delinquency bucket
dpd_bucket	string	N	Delinquency bucket	0|1-29|30-59|60-89|90+	Derived from dpd if provided
nonaccrual_flag	string	N	Nonaccrual indicator	Y|N	
risk_grade	string	N	Internal risk grade / rating	Reference table	Provide grade as of as_of_date
fico_score	int	N	Bureau score	300-850	If multiple scores, provide the primary score
utilization	decimal	N	Balance / credit line	0-1	Derived if credit_line provided

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
chargeoff_flag	string	N	Charge-off indicator	Y|N	Should align with CHARGEOFF table
days_on_books	int	N	Months since origination (approx)	>=0	Derived from origination_date if omitted
modification_flag	string	N	Modification / workout flag	Y|N	Terminology varies by client policy
forbearance_flag	string	N	Forbearance indicator	Y|N	
segment_override	string	N	Optional segment override code	Reference table	Used only if configured

5. TRANSACTION

Transactions used for payment behavior, balance movement diagnostics, and (optional) prepayment estimation.

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
txn_id	string	Y	Transaction identifier	<=64 chars	
account_id	string	Y	Account identifier		
post_date	date	Y	Posting date	YYYY-MM-DD	
effective_date	date	N	Effective date	YYYY-MM-DD	If differs from post_date

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
txn_type	string	Y	Transaction type	PAYMENT| FEE| INTEREST| ADVANCE| CURTAILME NT	
amount	decimal	Y	Transaction amount	Signed decimal	Payments negative, advances positive
principal_am ount	decimal	N	Principal component	Signed decimal	Optional
interest_amo unt	decimal	N	Interest component	Signed decimal	
fee_amount	decimal	N	Fee component	Signed decimal	
reversal_flag	string	N	Reversal indicator	Y|N	If Y, should negate original txn
source_syste m	string	Y	Source system code	Reference	

5. CHARGEOFF

Charge-off events used to define default outcomes and loss severity.

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
chargeoff_id	string	Y	Charge-off event id		
account_id	string	Y	Account identifier		
chargeoff_dat e	date	Y	Charge-off date	YYYY-MM-DD	

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
chargeoff_amount	decimal	Y	Charge-off amount	>=0	Provide gross charge-off; recoveries in RECOVERY
principal_chargeoff	decimal	N	Principal portion	>=0	
interest_chargeoff	decimal	N	Interest portion	>=0	
reason_code	string	N	Charge-off reason	Reference	
source_system	string	Y	Source system code	Reference	

5. RECOVERY

Recoveries associated with charged-off accounts; used to model recovery timing and net severity.

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
recovery_id	string	Y	Recovery event id		
account_id	string	Y	Account identifier		
recovery_date	date	Y	Recovery posting date	YYYY-MM-DD	
recovery_amount	decimal	Y	Recovery cashflow amount	>=0	Include liquidation proceeds as recoveries
recovery_type	string	N	Recovery type	CASH| LIQUIDATION| INSURANCE| OTHER	

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
cost_amount	decimal	N	Direct cost amount	>=0	Optional; used to net recoveries
collateral_id	string	N	Collateral identifier		If linked to collateral liquidation
source_system	string	Y	Source system code	Reference	

5. COLLATERAL

Collateral attributes and valuations (secured products).

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
collateral_id	string	Y	Collateral identifier		
account_id	string	N	Account identifier		If one-to-one; otherwise use bridge table
collateral_type	string	Y	Collateral type	AUTO|REAL_ESTATE|DEPOSIT|OTHER	
valuation_date	date	Y	Valuation date	YYYY-MM-DD	
collateral_value	decimal	Y	Collateral value	>=0	Appraised or estimated
valuation_type	string	N	Valuation method	APPRAISAL|AVM|GUIDEBOOK|OTHER	
lien_position	string	N	Lien position	1|2|3|OTHER	

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
zip_code	string	N	Collateral location	5-digit	
state	string	N	Collateral state	USPS 2-letter	

5. SCORE

Score observations used for segmentation and risk drivers.

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
score_id	string	Y	Score observation id		
customer_id	string	Y	Customer identifier		
account_id	string	N	Account identifier (if score tied to account)		
score_date	date	Y	Score date	YYYY-MM-DD	
score_type	string	Y	Score type	FICO|INTERNAL|BEHAVIORAL	
score_value	int	Y	Score value	Depends on type	FICO 300-850 typical
source_system	string	Y	Source system code	Reference	

5. MACRO_SERIES

Macroeconomic scenario series used during the forecast horizon and reversion period.

Field	Type	Req	Description	Allowed / Format	Notes / Derivation
scenario_set	string	Y	Scenario set identifier	<=64 chars	E.g., 2026Q1_ForecastSet_A
scenario_name	string	Y	Scenario name	BASELINE| ADVERSE| SEVERE| CUSTOM	
as_of_date	date	Y	Scenario vintage date	YYYY-MM-DD	The scenario package date
period_date	date	Y	Forecast period date	YYYY-MM-DD	Monthly or quarterly
series_name	string	Y	Macro series name	UNEMP| GDP|HPI| UVPI|RATE	Configured mapping
series_value	decimal	Y	Macro series value	Numeric	Unit conventions documented per series
units	string	N	Units	%|INDEX| RATE	

6. Mapping Guidance and Common Patterns

Mapping translates client source fields into the Northstar canonical schema. Northstar typically conducts an initial mapping workshop followed by iterative test loads. The mapping deliverable is maintained as a versioned artifact (mapping workbook) separate from this document.

6.1 Mapping Workflow (Recommended)

1. Step 1: Identify authoritative sources for each required field (system of record).
2. Step 2: Define transformation rules for each mapping (format, sign conventions, derivations).
3. Step 3: Load a pilot period (e.g., 3 months) to validate referential integrity and QC rules.
4. Step 4: Reconcile balances, charge-offs, recoveries to source totals; remediate gaps.
5. Step 5: Lock mapping version for production; route changes through change control.

6.2 Common Mapping Pitfalls

*   Delinquency definition mismatch (DPD vs billing cycle delinquency vs due-date delinquency).
*   Charge-off timing differences between servicing and GL posting date.
*   Recoveries netted vs gross; inclusion/exclusion of costs and fees.
*   Multiple collateral records per account without a consistent primary collateral rule.
*   Score recency and score selection (multiple bureaus/scores; missing score handling).
*   Account identifier changes due to system conversions or refinances.

6.3 Example Mapping Table (Excerpt)

Northstar Field	Client Source Field	Transform	Validation
ACCOUNT_SNAPSHOT.dpd	SERVICING.DAYS_PAST_DUE	CAST(int)	dpd >= 0; dpd_bucket derived
TRANSACTION.amount	CORE.PMT_AMT	Payments as negative	Sum(txn) ties to balance roll
CHARGEOFF.chargeoff_date	GL.POST_DATE	Use GL post date	Chargeoff_flag aligns to status_code
RECOVERY.recovery_amount	RECOV.CASH + RECOV.LIQUIDATION	Sum components	Non-negative; net costs if provided
COLLATERAL.collateral_value	APPRAISAL.VALUE	Latest by valuation_date	>= 0; stale > 18mo flagged

7. Transformations and Derivation Logic

This section outlines standard transformations applied during ETL. Clients may implement these transformations in their own pipelines or delegate to Northstar-managed ETL, subject to project scope.

7.1 Standardization Rules

- Trim whitespace; enforce UTF-8; reject non-printable characters in identifiers.
- Normalize missing values to empty string (flat files) or NULL (database).
- Enforce unique keys: (account_id, as_of_date) in ACCOUNT_SNAPSHOT.
- Amounts: use decimal with two to four decimal places; avoid scientific notation.

7.2 Delinquency Buckets

dpd_bucket = CASE WHEN dpd = 0 THEN '0' WHEN dpd BETWEEN 1 AND 29 THEN '1-29' WHEN dpd BETWEEN 30 AND 59 THEN '30-59' WHEN dpd BETWEEN 60 AND 89 THEN '60-89' ELSE '90+' END

7.3 Utilization (Revolving)

utilization = CASE WHEN credit_line > 0 THEN current_balance / credit_line ELSE NULL END

7.4 Net Recoveries (Optional)

net_recovery_amount = recovery_amount - COALESCE(cost_amount, 0)

7.5 Effective Interest Rate Proxy (If EIR Not Available)

If origination EIR is unavailable, clients may provide a product-level proxy rate. Proxy rates should be documented and reviewed under client governance. The proxy rate table is delivered as CONFIG_LOOKUP with effective dates.

7.6 Pseudocode: Monthly Snapshot Construction (Illustrative)

-- Build snapshot as-of date

INSERT INTO ACCOUNT_SNAPSHOT (account_id, as_of_date, current_balance, dpd, risk_grade, nonaccrual_flag, source_system)

SELECT a.account_id, :as_of_date, b.ending_bal, s.dpd, s.grade, s.nonaccrual, 'CORE'

FROM CORE_ACCOUNT a

JOIN CORE_BALANCES b ON a.account_id = b.account_id AND b.as_of_date = :as_of_date

LEFT JOIN SERVICING_STATUS s ON a.account_id = s.account_id AND s.as_of_date = :as_of_date;

8. Data Quality Controls and Rule Catalog

Northstar applies automated QC rules on each load. Rules are classified by severity (Critical, High, Medium, Low). Critical failures block processing; High failures require remediation or documented exceptions; Medium/Low are tracked.

8.1 QC Control Objectives

- Completeness: required fields present and populated above threshold.
- Validity: values within allowed ranges and formats.
- Consistency: cross-table alignment (e.g., charge-off flags vs charge-off events).
- Integrity: unique keys and referential integrity maintained.
- Reasonableness: statistical checks for outliers and drift vs prior periods.

8.2 Rule Catalog (Excerpt)

Rule ID	Table	Description	Severity	Threshold / Action
QC-001	ACCOUNT_SNAPSHOT	Missing account_id or as_of_date	Critical	0 allowed; reject file
QC-004	ACCOUNT_SNAPSHOT	Duplicate (account_id, as_of_date)	Critical	0 allowed; reject
QC-010	ACCOUNT_SNAPSHOT	dpd < 0 or dpd > 365	High	Investigate; cap only with exception
QC-021	CHARGEOFF	chargeoff_amount < 0	High	0 allowed; remediate
QC-034	RECOVERY	Recoveries posted before chargeoff_date	Medium	Allow if policy supports; document
QC-050	COLLATERAL	Valuation older than 18 months	Medium	Flag; consider conservative haircut
QC-072	ACCOUNT_SNAPSHOT	Balance drift > 20% vs prior month total	High	Reconcile; confirm portfolio event

8.3 Drift Monitoring (Data, Not Model)

Data drift is monitored using distribution shift metrics (e.g., PSI) for key drivers (score bands, utilization, risk grades) and volume checks (counts, balances). Drift alerts trigger investigation prior to model run sign-off.

9. Reconciliations and Control Totals

Reconciliations provide assurance that delivered data aligns to authoritative totals. Northstar recommends maintaining a monthly reconciliation package signed by data and finance owners.

9.1 Balance Reconciliation

- Tie total current_balance in ACCOUNT_SNAPSHOT to servicing or core month-end balances.
- Reconcile by major product/portfolio_code and by status_code (ACTIVE vs CLOSED).
- Explain differences (timing, excluded accounts, sold portfolios) and retain documentation.

9.2 Charge-off / Recovery Reconciliation

- Charge-offs: sum(chargeoff_amount) by month to GL charge-off accounts within tolerance.
- Recoveries: sum(recovery_amount) by month to GL recovery accounts; ensure sign conventions aligned.
- If recoveries are net of costs, document cost treatment and any excluded categories.

9.3 Tolerances (Typical)

Reconciliation	Typical Tolerance	Notes
Month-end balances	0.50% of total balance or $50k (whichever larger)	Higher tolerance may be required during conversions
Charge-offs	1.00% of monthly charge-offs	Must explain structural differences
Recoveries	2.00% of monthly recoveries	Recoveries often have timing differences

10. Security, Privacy, and Access Considerations

Client data may include PII. Northstar applies layered controls across transmission, storage, and access. The client remains responsible for limiting the data shared to what is necessary for CECL purposes.

10.1 PII and Sensitive Fields (Examples)
- Customer name
- SSN/TIN (not required; if provided, must be tokenized)
- Date of birth (not required)
- Address (optional; state/ZIP preferred)
- Account number (if used as account_id, should be hashed)

10.2 Encryption
- In transit: SFTP with key-based authentication; API with mutual TLS.
- At rest: AES-256 encryption in managed storage; keys managed by Northstar KMS.
- Field-level protection: optional hashing/tokenization for account_id and customer_id.

10.3 Access Model

Access is role-based and limited to implementation and support personnel with a defined business need. All data access is logged. Data retention defaults to 7 years unless otherwise agreed.

11. Change Management and Versioning

- Mapping versions are tracked with semantic versioning (e.g., MAP-1.3.0).
- Changes to required fields or sign conventions require testing in a non-production environment.
- Production changes require approvals from the client model owner and data owner.
- Northstar provides release notes for canonical schema updates; clients must assess downstream impacts.

11.1 Issue Handling and Defect Resolution

Data defects are triaged by severity. Critical defects block the run. A defect ticket should include sample records, source extracts, and reconciliation impacts. Once remediated, clients re-deliver the full affected table for the impacted as_of_date.

Appendix A: Sample File Excerpts (Illustrative)

A.1 ACCOUNT_SNAPSHOT (CSV)

account_id,as_of_date,current_balance,dpd,risk_grade,nonaccrual_flag,source_system

A100245,2026-01-31,12450.32,0,A1,N,CORE

A100871,2026-01-31,8023.10,35,B2,N,CORE

A101992,2026-01-31,0.00,0,A1,N,CORE

A.2 CHARGEOFF (CSV)

chargeoff_id,account_id,chargeoff_date,chargeoff_amount,reason_code,source_system

CO7781,A100871,2025-12-18,1543.22,LOSS,GL

CO7810,A104332,2026-01-09,312.10,LOSS,GL

Appendix B: Troubleshooting (Common Errors)

- Duplicate snapshot keys: verify account_id stability and remove duplicate extracts before delivery.
- Negative payments: confirm sign convention; payments should be negative amounts in TRANSACTION.
- Recoveries prior to charge-off: ensure recovery_date uses posting date; verify policy for pre-charge-off recoveries.
- Stale collateral: provide latest valuations or document conservative assumptions used for LGD.

Appendix C: Implementation Checklist (Excerpt)

6. Confirm authoritative source for each required table and field.
7. Complete mapping workbook and obtain approvals.
8. Deliver pilot loads for at least 3 consecutive periods.
9. Resolve all Critical and High QC findings.
10. Complete reconciliations to GL and servicing totals for pilot period.
11. Execute production dry-run and sign-off for go-live.