

## 1. Extraction Strategy Definition

Source Name	Extraction Mode (Pull / Push)	Processing Type (Batch / Incremental)	Load Strategy (Full / Delta / Snapshot)	Extraction Trigger	Notes
Salesforce / Edge	Pull	Incremental	Delta	Scheduled or Event	System integration, supports field such as 'Last Modified Timestamp' for change tracking.
Formstack	Pull	Incremental	Delta	Submission Event	Platform web forms, schema varies per form/template.
Schwab Custodial Platform	Pull	Batch	Snapshot	Account Opening Event	Platform integration, structure may change per custodian/rule update.
Fidelity Custodial Platform	Pull	Batch	Snapshot	Account Opening Event	As above.
DocuSign	Pull	Incremental	Delta	Envelope Event	Envelope/document routing, schema varies, triggered per document transaction.
Black Diamond (optional)	Pull	Batch	Snapshot	As Needed	Optional retrieval for verification; presence/value is conditional.
Client-provided Statements/Docs	Push	Event-driven	Snapshot	Client Upload	Unstructured (PDF, etc.), uploaded via Formstack or other; variable contents.
Subatomic Deep Lens	Pull	Incremental	Delta	Audit Event	Platform/module, logs generated per workflow event.

## 2. ETL vs ELT Strategy Determination

Source Name	Strategy Type (ETL / ELT)	Justification Based on Source Constraints	Pre-Load Transformations (If Any)	Post-Load Transformations Expected	Notes
Salesforce / Edge	ELT	Structure is stable and suitable for raw preservation, all fields to be retained as-is.	None	Schema adjustments in downstream	No decoding/format change needed.
Formstack	ELT	Schema varies, but raw form and fields must be preserved; normalization deferred.	None	Field mapping, completeness checks	No pre-load changes allowed due to high variability and template evolution.
Schwab Custodial Platform	ELT	Custodian/rule changes expected; preserving as-is supports traceability.	None	Validation, field/rule alignment	Only custodial format normalization might be needed post-load.
Fidelity Custodial Platform	ELT	Same as above.	None	Validation, field/rule alignment	No mandatory pre-load transformation requirement.
DocuSign	ELT	Envelope data/fields are highly variable; must persist raw content.	None	Envelope audit, normalization	No decoding or format change required for extract.
Black Diamond (optional)	ELT	Statement format varies by provider; retain original PDF/raw artifact.	None	Field extraction/verification	Only needed when provided.
Client-provided Statements/Docs	ELT	Structure unpredictable; raw documents must be retained for audit/compliance.	None	Parsing/classification as needed	No pre-load normalization due to unstructured nature.
Subatomic Deep Lens	ELT	Audit/event logs in system-native, possible nested form. No transformation needed pre-load.	None	Trace correlation, scoring, lineage	Preserve all audit events at their native granularity.

## 3. Extraction Pattern Identification

Source Name	Pattern Type (Date / Identifier / Snapshot)	Change Detection Mechanism	State Required (Yes / No)	State Scope (Local / External)	Notes
Salesforce / Edge	Date/Identifier	Last Modified Timestamp / unique IDs	Yes	Local	For incremental extraction (delta) based on modification.
Formstack	Identifier	Submission Event/ Unique Submission ID	Yes	Local	Each form submission is discrete.
Schwab Custodial Platform	Snapshot	Full set at event time (e.g., account open)	No	N/A	Full library/rule snapshot at account opening.

Source Name	Pattern Type (Date / Identifier / Snapshot)	Change Detection Mechanism	State Required (Yes / No)	State Scope (Local / External)	Notes
Fidelity Custodial Platform	Snapshot	As above	No	N/A	Same as Schwab.
DocuSign	Identifier	Envelope/Document ID, envelope events	Yes	Local	Each event/envelope uniquely identified.
Black Diamond (optional)	Snapshot	Optional/Conditional Snapshot	No	N/A	Statement retrieval only when triggered.
Client-provided Statements/Docs	Identifier	Upload Event/File Identifier	No	N/A	Each document is new, uniquely referenced by upload.
Subatomic Deep Lens	Date/Identifier	Audit Log Event/Trace Timestamp	Yes	Local	Each audit event linkable by collection timestamp/trace.

#### 4. Dependency and Availability Analysis

Source Name	Frequency	Availability Window	Ordering Constraints	Dependency Type (None / Temporal / External)	Notes
Salesforce / Edge	Scheduled/Event	High (system of record)	None	None	Can extract at any point, independent.
Formstack	Event-driven	On submission	None	None	Each submission is atomic; no dependency.
Schwab Custodial Platform	As triggered	On account open	None	None	No inter-source dependencies.
Fidelity Custodial Platform	As triggered	On account open	None	None	Same as Schwab.
DocuSign	Per document	On envelope creation	None	None	No ordering dependency; each envelope independent.
Black Diamond (optional)	On demand	As needed	None	None	Only triggered when needed.
Client-provided Statements/Docs	Event-driven	On upload	None	None	Ingestion per document, no dependency.
Subatomic Deep Lens	Continuous/Batch	Per audit event	None	None	Audit logs consumed as available.

#### 5. Versioning and Replay Strategy

Source Name	Versioning Method	Version Granularity	Replay Supported (Yes / No)	Replay Scope	Notes
Salesforce / Edge	Timestamp/Identifier	Record (per change)	Yes	By record/date	Can reconstruct state as of timestamp or last modified.
Formstack	Submission ID/Datetime	Submission	Yes	By submission	Replay possible per submission event.
Schwab Custodial Platform	Datetime/Snapshot ID	Library/rule snapshot	Yes	By snapshot	Full snapshot can be replayed at point in time.
Fidelity Custodial Platform	Datetime/Snapshot ID	As above	Yes	By snapshot	Same as Schwab.
DocuSign	Envelope/Event ID	Envelope	Yes	By envelope	Each envelope retrievable individually for replay.
Black Diamond (optional)	Retrieval Ts/Doc ID	Statement PDF	Yes	By statement	Each statement stored with retrieval timestamp.
Client-provided Statements/Docs	Upload ID/Timestamp	Document	Yes	By document	Original item locally versioned by upload event.
Subatomic Deep Lens	Trace/Event ID, Timestamp	Event/log trace	Yes	By trace/event	Audit trails fully versioned by event time.

#### 6. Reliability and Fault Isolation Rules

Source Name	Retry Strategy	Idempotency Mechanism	Error Classification	Failure Isolation Level	Notes
Salesforce / Edge	Retry on failure	Record/IDempotent extraction	Extraction, connection	Source-level	Retries on API/system failure, safe re-reads.

Source Name	Retry Strategy	Idempotency Mechanism	Error Classification	Failure Isolation Level	Notes
Formstack	Retry on failure	Submission/Event ID	Extraction	Source-level	Ingestion per form, duplicates avoided by ID.
Schwab Custodial Platform	Retry on failure	Snapshot/Datetime	Extraction	Source-level	Snapshots can be retried without cross-source issue.
Fidelity Custodial Platform	Retry on failure	Snapshot/Datetime	Extraction	Source-level	Same as Schwab.
DocuSign	Retry on failure	Envelope/Event ID	Extraction	Source-level	Envelope IDs prevent duplicates during retry.
Black Diamond (optional)	Retry on error	Statement/Doc ID	Extraction	Source-level	Each optional retrieval isolated to source/process.
Client-provided Statements/Docs	N/A (manual)	Upload/Event ID	Extraction	Item-level	Isolation per document upload.
Subatomic Deep Lens	Retry on error	Trace/Event ID	Extraction/log event	Source-level	Idempotent, log replay or error recovery per event.

## 7. Orchestration Readiness (Decoupled View)

Source Name	Trigger Type (Time / Event / Manual)	Async / Sync	Requires Queue or Buffer (Yes / No)	Orchestrator Compatibility Notes
Salesforce / Edge	Time/Event	Async	No	Supports decoupled triggers, event or schedule.
Formstack	Event	Async	Yes	Event trigger (form submit); may queue high bursts.
Schwab Custodial Platform	Event	Async	No	Triggered by account opening workflow.
Fidelity Custodial Platform	Event	Async	No	Same as Schwab.
DocuSign	Event	Async	Yes	Queue for high envelope volume possible.
Black Diamond (optional)	Manual/Event	Async	No	Only initiated on demand.
Client-provided Statements/Docs	Event/Manual	Async	No	Per upload; event triggers extraction.
Subatomic Deep Lens	Event/Time	Async	Yes	Log events may be buffered during burst.

## 8. Infrastructure Implications per Source

Source Name	Temporary Storage Needed (Yes / No)	Compute Intensity (Low / Medium / High)	Expected Data Volume	Cost Sensitivity Drivers	Notes
Salesforce / Edge	No	Low	Moderate	API usage, frequency	Structured, predictable system data intake.
Formstack	Yes	Medium	Variable, could spike	Submission rate, schema drift	Needs queue/buffer for burst, variety in payloads.
Schwab Custodial Platform	No	Low	Low to moderate	Snapshot size	Full sets at key events.
Fidelity Custodial Platform	No	Low	Low to moderate	Snapshot size	Same as above.
DocuSign	Yes	Medium	Low to moderate, event-based	Envelope volume	Temporary storage for envelope/artifact.
Black Diamond (optional)	No	Low	Low (optional)	Statement count	Only as needed.
Client-provided Statements/Docs	Yes	Medium	Potentially high, unstructured	Storage for PDFs/docs	Unstructured data can consume significant space.
Subatomic Deep Lens	Yes	Medium	Moderate (log/audit)	Audit depth, retention	Temporary buffer for high-frequency logs.