

## Silver Layer Rules Extraction

### Entity, Attribute, and Relationship Specification

Entity Name	Attributes	Relationships	Justification
IntakeFormNormalized	formid (string), templatename (string), clientid (string, nullable), householdid (string, nullable), submissiontimestamp (datetime), field[X] (various, flattened), sourcefile (string), ingestontimestamp (datetime)	1:N DocumentMetadata (link on clientid or formid), 1:N EnvelopeMetadata (through doclink or embedded reference), 1:N AuditEventLog (by formid)	Required to represent each normalized intake form submission with all relevant metadata, field values, and linkage to client/household, supporting validation, packet assembly, exception review, and audit trace; forms drive workflow initiation and are reference points for further artifacts and auditability.
Document_Metadata	documentid (string), clientid (string, nullable), documenttype (string), uploadtimestamp (datetime), filehash (string), sourcesystem (string)	N:1 IntakeFormNormalized (via clientid or referenced formid), N:1 EnvelopeMetadata (via doclink), 1:N AuditEventLog (by document_id, if referenced)	Needed to track and deduplicate every client-provided document (statements, IDs, etc.), associate uploads to clients and forms, enable metadata-driven packet assembly, and provide document lineage for audit/compliance.
CustodianFormIndex	custodian (string), formid (string), accounttype (string, nullable), version (string, nullable), ingestontimestamp (datetime), filename (string), source_system (string)	None direct; referenced by IntakeFormNormalized during packet assembly	Holds all available custodian form templates and associated metadata; enables correct, rule-driven selection and population of custodian forms based on intake and validation logic; required for compliance and document completeness checks.
Envelope_Metadata	envelopeid (string), eventtimestamp (datetime), status (enum: sent/completed/cancelled/error), recipientemail (string, nullable), recipientrole (string, nullable), doclink (string, nullable), sourcesystem (string)	N:1 IntakeFormNormalized (via doclink or embedded form reference), N:1 DocumentMetadata (via doclink), 1:N AuditEventLog (by envelopeid, if referenced)	Represents each DocuSign envelope event/status necessary for tracking signature flows, legal completion, and coordinating form packet workflow; must support audit, traceability, and recovery in exception cases.
AuditEventLog	auditeventid (string), eventtype (string), actorid (string, nullable), eventtimestamp (datetime), relatedformid (string, nullable), sourcesystem (string)	N:1 IntakeFormNormalized (via relatedformid), N:1 DocumentMetadata (via documentid), N:1 EnvelopeMetadata (via envelopeid)	Captures normalized, traceable audit logs and workflow events for every relevant step and change; required for compliance, operational review, and end-to-end traceability—ensuring Silver Layer data always links to workflow events and decisions.
FormDocumentAssoc	formid (string), documentid (string)	Many-to-many bridge: IntakeFormNormalized ↔ Document_Metadata	Where direct references aren't normalized in Document_Metadata, a bridge is required to link intake forms to referenced/embedded documents, ensuring all document associations are tracked for packet assembly, review, and compliance.
EnvelopeDocumentAssoc	envelopeid (string), documentid (string)	Many-to-many bridge: EnvelopeMetadata ↔ DocumentMetadata	Needed to associate DocuSign envelopes to specific documents (such as completed forms or attached IDs) whenever those links aren't 1:1 or are not otherwise explicit in the base tables; required for completeness and traceability in the digital signature/review process.

### Business Rules

Rule Description	Source Dataset	Target Dataset	Priority	Notes
All required fields for IntakeFormNormalized (formid, templatename, submissiontimestamp, sourcefile) must be present and non-null	RawIntakeForms	IntakeFormNormalized	High	Enforced at ingestion to Silver; incomplete rows are rejected or quarantined
All date/time fields (e.g., submissiontimestamp, ingestontimestamp, eventtimestamp, uploadtimestamp) must be converted to ISO 8601 string format	All Bronze datasets	All Silver tables	High	Ensures downstream alignment and time-based logic; applies Silver-wide
Each IntakeFormNormalized row must be linked via clientid or formid to at least one valid Document_Metadata row for supporting documents, if referenced	RawIntakeForms, RawClientDocuments	IntakeFormNormalized, Document_Metadata	High	Required for audit traceability and validation completeness
Only one copy of each uploaded document (client statement, ID) with a unique filehash and documentid per upload/session may be included in Document_Metadata	RawClientDocuments, Raw_Statements	Document_Metadata	High	Deduplication rule critical for audit, prevents double-counting of documents
Envelope_Metadata.status must be one of the standardized enum values {sent, completed, cancelled, error}	RawDocuSignEnvelopes	Envelope_Metadata	High	Downstream logic depends on reliable, normalized status values

Rule Description	Source Dataset	Target Dataset	Priority	Notes
AuditEventLog must contain auditeventid, eventtimestamp, sourcesystem at a minimum for every record	RawAuditEvent_Logs	AuditEventLog	High	Required to guarantee full traceability/compliance; nonconforming records are flagged
CustodianFormIndex rows must have custodian and form_id populated (version is nullable); incomplete records are dropped or flagged	RawCustodianForm_Libraries	CustodianFormIndex	High	Required for valid packet assembly and form lookup
IntakeFormNormalized must flatten all nested fields in \$.fields and align naming to agreed schema; any fields not mapped are placed in reserved columns	RawIntakeForms	IntakeFormNormalized	Medium	Ensures every form is parseable and compatible for workflow processing
DocumentMetadata.documenttype must be extracted reliably to (statement, ID, etc.); ambiguous types are set to 'unknown' and may require manual review	RawClientDocuments, Raw_Statements	Document_Metadata	Medium	Ensures downstream packet assembly and eligibility checks
Each Silver row must have ingestion_timestamp consistently populated from the Bronze ingest timestamp metadata	All Bronze datasets	All Silver tables	High	Enables incremental, replayable, and auditable processing
Only records originating from approved Bronze datasets/sources (Formstack, DocuSign, Schwab/Fidelity, Client Upload, Subatomic Deep Lens) are included in Silver	All Bronze datasets	All Silver tables	High	Fulfils IT, security, and audit requirements: data provenance/enforcement
EnvelopeMetadata rows must unnest recipient lists so each recipient_email and recipient_role appears as atomic columns (or rows if multi-value normalized)	RawDocuSignEnvelopes	Envelope_Metadata	Medium	Needed for routing, review, and downstream signature tracking

## Data Transformation Rules

Rule Description	Source Field	Target Field	Transformation Type	Notes
Map and flatten nested form fields from RawIntakeForms (\$.fields) into individual IntakeFormNormalized attributes	\$.fields	field_[X] (flattened columns)	Flatten/Normalization	Handles variable intake schemas across different templates
Parse and align source formid from \$.formid or \$.eventid to IntakeFormNormalized.formid	\$.formid / \$.eventid	form_id	Mapping	Ensures consistent primary key for linking and traceability
Extract and conform clientid and householdid from form payload if present	\$.clientid, \$.householdid	clientid, householdid	Mapping	Optional, but aligned to operational use cases
Convert all timestamp values (submission, upload, ingestion, event) to ISO 8601 string format	\$.submissiontime, uploadts, \$.event_time, etc.	submissiontimestamp, uploadtimestamp, ingestiontimestamp, eventtimestamp	Format Alignment/Conversion	Entire Silver Layer; ensures temporal consistency
Extract document attributes (documentid from filename or fileid, filehash, documenttype) and standardize	filename, fileid, filehash, mime, extracted type	documentid, filehash, document_type	Conform/Extraction/Mapping	Deduplication and semantic alignment
Deduplicate documents in DocumentMetadata by filehash and (filename, upload_timestamp) composite key	filename, filehash, uploadts	documentid, filehash, upload_timestamp	Deduplication/Conform	Ensures one row per unique document/upload event
Extract custodian name, formid, accounttype, and version from file properties and path in RawCustodianForm_Libraries	path, filename, embedded metadata	custodian, formid, accounttype, version	Conform/Extraction/Mapping	Enables rules-based form packet selection
Unnest and map envelope metadata and attached document references in RawDocuSignEnvelopes	\$.recipients[], \$.documents[]	recipient_email, recipientrole, doc_link	Flatten/Unnest/Mapping	Ensures recipients and document linkage are queryable and reviewable
Normalize Envelope_Metadata.status field to enum {sent, completed, cancelled, error}	\$.status	status	Format/Stabilization	Bridges inconsistency in underlying workflow sources
Map audit log event identifiers and timestamps, even if missing or variable in source, to canonical field names	\$.eventid, \$.logid, \$.event_time, \$.timestamp	auditeventid, event_timestamp	Alignment/Mapping	Required for unambiguous trace, deduplication, and ordering
Create association rows in FormDocumentAssoc and EnvelopeDocumentAssoc where document references are present	Embedded or referenced filenames/IDs	formid/documentid, envelopeid/documentid	Extraction/Bridge Table Mapping	Needed when links not explicit in base tables
Retain ingestion_timestamp metadata from all raw Bronze objects for each silver row	Bronze-level ingest metadata	ingestion_timestamp	Mapping/Propagation	

Rule Description	Source Field	Target Field	Transformation Type	Notes
				For auditability, reprocessing, and lineage