

FileMetadata Table Schema FieldSchema Table Schema FieldConfidence Table Schema Description Attribute Attribute Unique identifier for the Description Description FileId (PK) file FieldID (PK) Unique identifier for the field FieldID (PK) Unique identifier for the field Batch identifier for grouping files Process type FileID Associated file identifier ProcessType (Transfection, etc) Date the file was UploadDate uploaded BatchID Batch identifier for grouping fields FieldDescription Description of the field Process type ProcessType FieldName (Transfection, etc) Name of the field **ValidationRules** Rules for validation ExtractedValue Value extracted by Textract Status Status of the file DateAdded Date the field was added Error details ConfidenceScore Confidence score from Textract

Indicates if

reprocessing is needed

Path to the file in S3

LowConfidenceFlag | True if below confidence threshold

Metadata schemas

Reprocess Flag

S3 loc

Job Name	Purpose	Job Name	Purpose
validation_and_transform_job.py	Validates and transforms data	file_metadata_handler.py	Handles file uploads and logs metadata
batch_aggregator_job.py	Merges data for a BatchID	textract_processor.py	Processes files using Textract
impala_loader_job.py	Loads aggregated data into Impala	confidence_logger.py	Maps fields and evaluates confidence
		metadata_updater.py	Updates file metadata across all steps