

Technical Description: Lambda Logic & Control Flow

1. Submission Lambda Flow

Trigger: S3 ObjectCreated (incoming/*.pdf)

Steps:

1. Parse Bucket and Key from event.
2. Validate file extension.
3. Call Textract.StartDocumentAnalysis(NotificationChannel=SNS).
4. Log JobId to CloudWatch.
5. Exit (Stateless).

2. Processing Lambda Flow

Trigger: SQS Message (Batch)

Steps:

1. Parse Message Body -> SNS -> JobId & Status.
2. If Status != SUCCEEDED -> Move to error/ -> Return.
3. Call Textract.GetDocumentAnalysis(JobId).
4. If NextToken exists -> Loop until all blocks retrieved.
5. Parse Blocks -> Filter Confidence > 90% -> Extract Fields.
6. Generate CSV -> S3 PutObject (processed/).
7. Copy Source -> S3 (archive/).
8. Delete Source -> S3 (incoming/).

3. Error Handling & DLQ

- Try/Except blocks wrap all AWS SDK calls.
- SQS Visibility Timeout ensures retries on transient failures.
- Dead Letter Queue (DLQ) captures messages after 3 failed attempts.