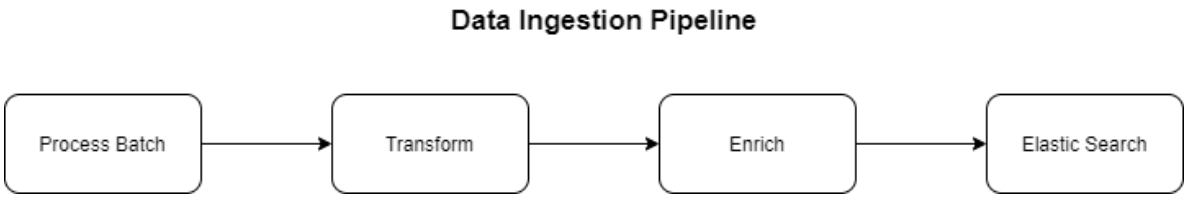
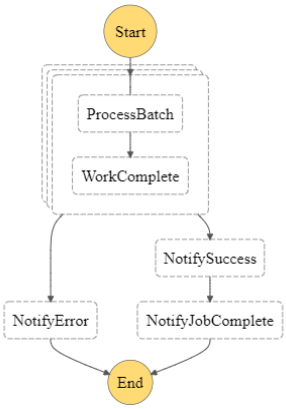
Cost Estimations

Following document illustrates the cost involved in the data-ingestion, components of the pipeline are described in the diagram below:

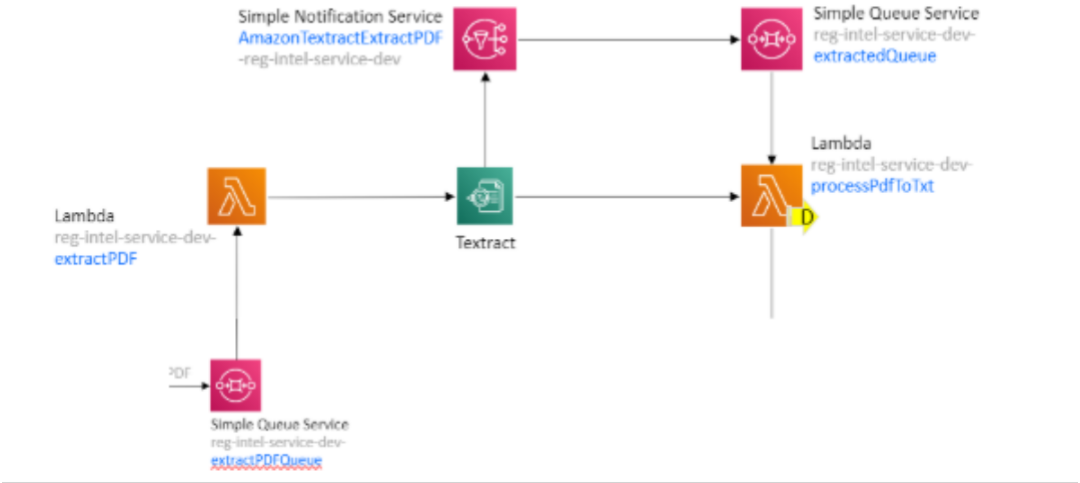


1. Process Batch



Functions	Memory	Time	Description
load_parameters (**lambda**)	2240MB	15min (900 sec)	statemachine invoked at the end of every 100 files in 2 chunks
regques_edb_StateMachine (**sf**)			invocations
process_batch (**lambda**)	3008MB	15min (900)	save events into s3
notify complete (lambda)	128	15min (900)	send sns when job completes

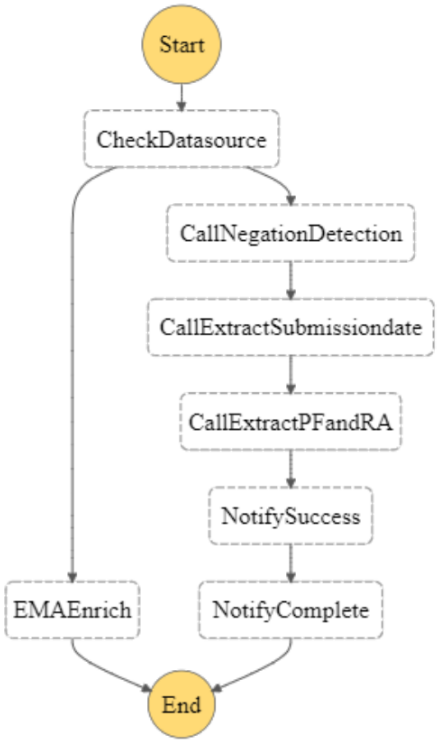
2. Transform



Components

Function/Service	Memory (MB)	Time	Description
extractPDF	2240	15min or 900 sec	lambda is triggered during s3 events
regquest_edb_transform_pdfTotxt	256 KB (max message size)	20 min	sqs queue
textract			AWS textract
sns notification			sns notification
regquest_edb_pdftotxt_lambda	2240	15min or 900	lambda function triggered on completion of textract
edb_regquest_transform_dynamodb			dynamodb table

3. Enrich



Function/Service	Memory	Time	Description
regquest_gra_enrich_invoke_sm_lambda (lambda)	2240	15min	Lambda triggered when there is s3 event
regquest_enrich_StateMachine (statemachine)			
regquest_gra_enrich_nd_lambda	2240	15min	
regquest_gra_enrich_sd_lambda	2240	15min	
regquest_gra_enrich_pfRa_lambda	2240	15min	
regquest_gra_enrich_jc_lambda	128	15min	

4. Elastic Search

Function/Service	Memory	Time	Description
regquest_ES_DataIngestion_lambda	128	30sec	Elastic ingestion lambda