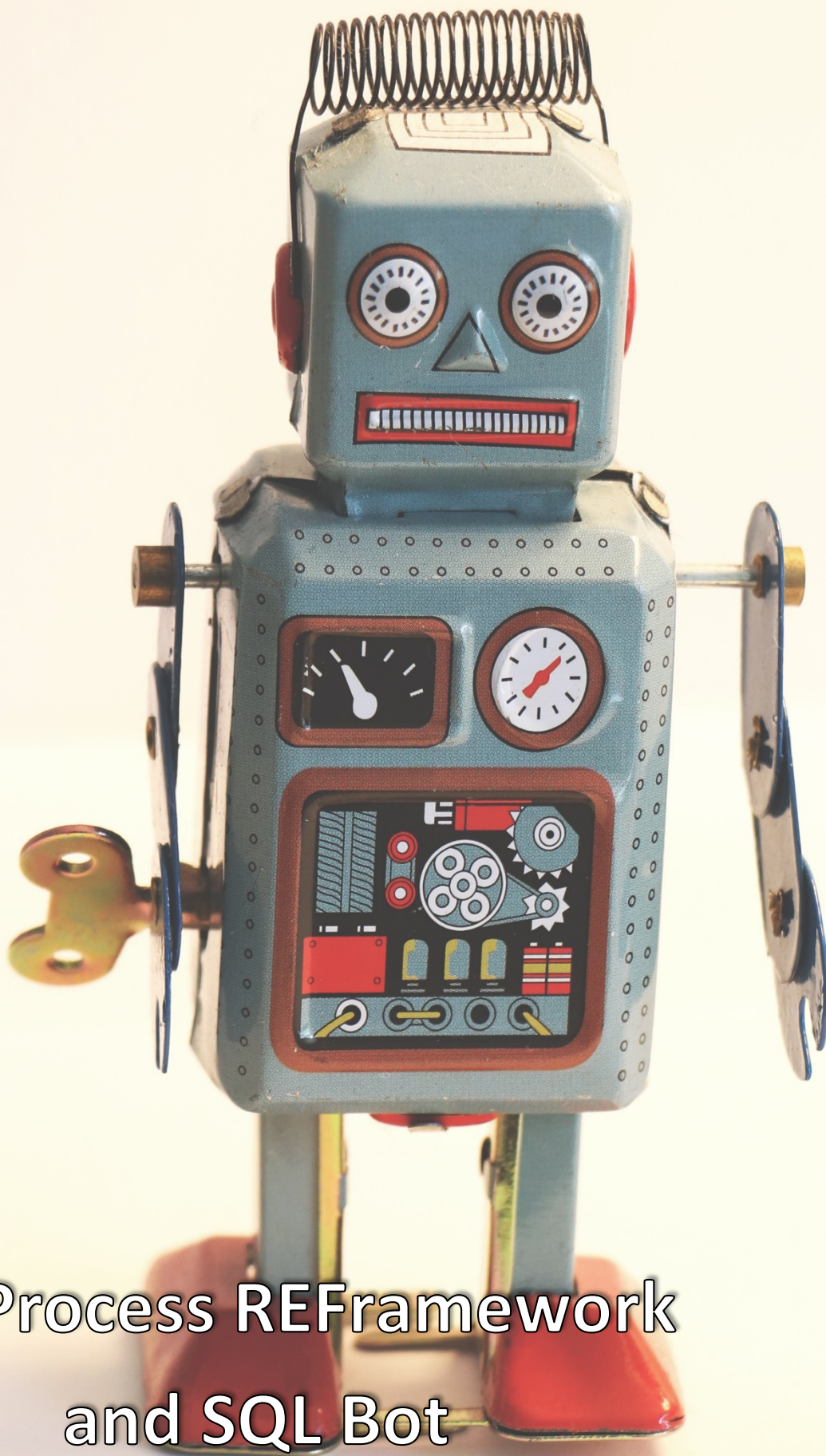


Project by Brayan Jimenez



Doc Process REFramework  
and SQL Bot

As is process:

Process of video:

<https://drive.google.com/file/d/1DSfbOZSYEUUyiYlwsXX0dLP8wT3VoD3f/view?usp=sharing>

A company receive the following types of documents via Outlook:

- Invoice
- Receipt
- W9 Form
- Certificate of filing

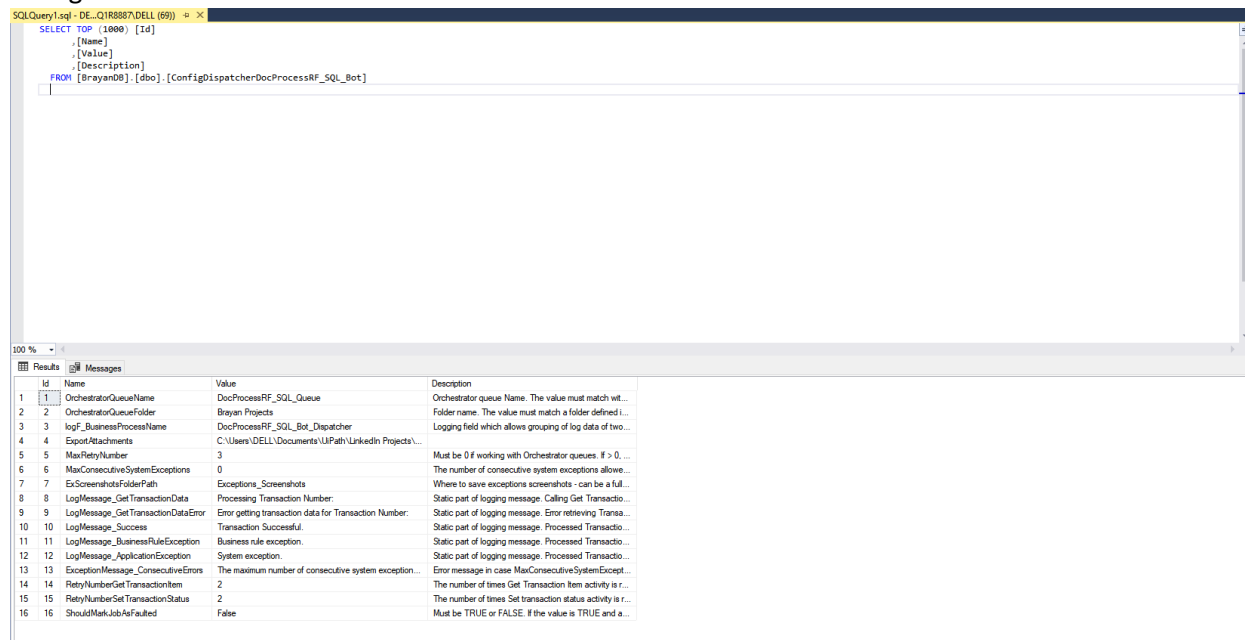
These documents are required to be extracted from the email automatically, then downloaded, then uploaded to rows in the orchestrator, then processed to extract the data from each type of document for later export to Excel, the information will be extracted from each pdf file and then exported in an Excel for each type of document.

Triggers should be used, to trigger the dispatcher, and triggers to trigger the performer.

To be process:

#### Dispatcher:

- It's mandatory to use database for this automation, in this case in order to stored the config file



```

SELECT TOP (1000) [Id]
, [Name]
, [Value]
, [Description]
FROM [BrayanDB].[dbo].[ConfigDispatcherDocProcessRF_SQL_Bot]

```

Id	Name	Value	Description
1	OrchestratorQueueName	DocProcessRF_SQL_Queue	Orchestrator queue Name. The value must match wit...
2	OrchestratorQueueFolder	Brayan Projects	Folder name. The value must match a folder defined i...
3	logF_BusinessProcessName	DocProcessRF_SQL_Bot_Dispatcher	Logging field which allows grouping of log data of two...
4	ExportAttachments	C:\Users\DELL\Documents\UIPath\LinkedIn Projects\...	
5	MaxRetryNumber	3	Must be 0 if working with Orchestrator queues. If > 0, ...
6	MaxConsecutiveSystemExceptions	0	The number of consecutive system exceptions allowe...
7	ExScreenshotsFolderPath	Exceptions_Screenshots	Where to save exceptions screenshots - can be a full...
8	LogMessage_GetTransactionData	Processing Transaction Number:	Static part of logging message. Calling Get Transac...
9	LogMessage_GetTransactionDataError	Error getting transaction data for Transaction Number:	Static part of logging message. Error retrieving Transa...
10	LogMessage_Success	Transaction Successful.	Static part of logging message. Processed Transac...
11	LogMessage_BusinessRuleException	Business rule exception.	Static part of logging message. Processed Transac...
12	LogMessage_ApplicationException	System exception.	Static part of logging message. Processed Transac...
13	ExceptionMessage_ConsecutiveErrors	The maximum number of consecutive system exception...	Error message in case MaxConsecutiveSystemExcept...
14	RetryNumberGetTransactionItem	2	The number of times Get Transaction Item activity is r...
15	RetryNumberSetTransactionStatus	2	The number of times Set transaction status activity is r...
16	ShouldMarkJobAsFailed	False	Must be TRUE or FALSE. If the value is TRUE and a...

- If the table isn't created it'll be used a workflow called "Load Config File to SQL" and this will take care of uploading the config to the database in addition to creating the table
- A queue is created in the orchestrator which is where all the documents will be loaded afterwards.
- The dispatcher is created by means of RFramework for linear process to upload the documents downloaded via Outlook to the Queue in the orchestrator.
- Integration service is used to use an event trigger through a connection with Outlook 365. This trigger will be used by the dispatcher and will allow the dispatcher to be automatically triggered when new items, in this case downloaded documents, are

uploaded to the queue. The event trigger monitors Outlook to receive an email with the subject "Uipath" and containing documents.

### Performer:

- The performer is created by means of the Document Understanding Framework and the file is configured, first the settings are configured and the DocumentUnderstandingQueueName, Target file, ExportsFolder, etc. are indicated. Then the assets sheet is configured. For this, all the assets are first created in the orchestrator to be able to use them in the asset sheet of the config file.
- Here also it'll be used database table to store the config file of performer

```
SQLQuery2.sql - DE_Q1R8887DELL (79) - X
SELECT TOP (1000) [Id]
, [Name]
, [Value]
, [Description]
, [NameAsset]
, [Asset]
, [DescriptionAsset]
FROM [Brayan08].[dbo].[ConfigPerformerDocProcessRF_SQL_Bot]
```

Results							
Messages							
Id	Name	Value	Description	NameAsset	Asset	DescriptionAsset	
1	TemporaryLocalFolder	Data\TempFolder	Auxiliary folder where the job will download files wh...	ApiKey	DocumentUnderstandingApiKey	Retrieve your DocumentUnderstanding ApiKey and c...	
2	ExportsFolder	C:\Users\DELL\Documents\UPath\Linked Projects\...	Output folder for exported data	AlwaysValidateClassification	AlwaysValidateClassificationDP_RFSQL_Bot	If set to True, classification will always go through ma...	
3	ReceiptsTrainingFolder	Data\ReceiptsTraining\	Path for ML model training - Receipts	AlwaysValidateExtraction	AlwaysValidateExtractionDP_RFSQL_Bot	If set to True, extracted data will always go through m...	
4	InvoicesTrainingFolder	Data\InvoicesTraining\	Path for ML model training - Invoices	SkipClassifierTraining	SkipClassifierTrainingDP_RFSQL_Bot	If set to True, classifier training will not be performed...	
5	ClassifierLearningFilePath	DocumentProcessing\IntelligentKeywordLearningFile.json	Path to the learning file of the classifier	SkipExtractorTraining	SkipExtractorTrainingDP_RFSQL_Bot	If set to True, extractor training will not be performed...	
6	UPathOCREndpoint	https://du.upath.com/ocr	UPath public endpoint for DocumentOCR	ActionCatalog	ActionCatalog_DP_RFSQL_Bot	Action Catalog Name. The asset value overrides the...	
7	ReceiptsEndpoint	https://du.upath.com/ie/receipts	UPath public endpoint for receipts	ActionFolderPath	ActionFolderPath_DP_RFSQL_Bot	Path to the Orchestrator Folder where the Action Cata...	
8	InvoicesEndpoint	https://du.upath.com/ie/invoices	UPath public endpoint for invoices	StorageBucketName	StorageBucket_DP_RFSQL_Bot	Storage Bucket Name (required when Action Center i...	
9	ClassificationEndpoint	https://du.upath.com/svc/intelligentkeywords	UPath public endpoint for classification	StorageBucketDirectoryPath	StorageBucketPath_DP_RFSQL_Bot	Path inside the Storage Bucket where actions will sto...	
10	FormExtractorEndpoint	https://du.upath.com/svc/formextractor	UPath public endpoint for Form Extractor	DocumentUnderstandingQueueName	QueueName_DP_RFSQL_Bot	The name of the Orchestrator Queue. The asset valu...	
11	AlwaysValidateClassification	True	If set to True, classification will always go through ...	DocumentUnderstandingQueuePath	QueuePath_DP_RFSQL_Bot	Path to the Orchestrator Folder where the Queue resi...	
12	AlwaysValidateExtraction	True	If set to True, extracted data will always go through...	ClassificationThreshold	ClassificationThreshold_DP_RFSQL_Bot	Classification threshold for Classification Business Rul...	
13	SkipClassifierTraining	False	If set to True, classifier training will not be performe...				
14	SkipExtractorTraining	False	If set to True, extractor training will not be performe...				
15	ClassificationThreshold	0.5	Classification threshold for Classification Business R...				
16	TargetFileKey	Path	Key in the Transaction Item's dictionary that holds t...				
17	ActionCatalog		Action Catalog Name. Has a corresponding asset t...				
18	ActionFolderPath		Path to the Orchestrator Folder where the Action C...				
19	StorageBucketName		Storage Bucket Name (required when Action Cente...				
20	StorageBucketDirectoryPath		Path inside the Storage Bucket where actions will ...				
21	DocumentUnderstandingQ...	DU_Invoice	The name of the Orchestrator Queue. Has a corres...				
22	DocumentUnderstandingQ...		Path to the Orchestrator Folder where the Queue re...				
23	MaxExecutionAttempts	3					
24	MaxExecutionAttemptsDig...	3					
25	MaxExecutionAttemptsCla...	3					
26	MaxExecutionAttemptsExtr...	3					
27	MaxLockTimeout	30					
28	RetryInterval	30					
29	ClassificationActionTitle	Classify					
30	ValidationActionTitle	Validate					
31	LogMessage_DocumentT...	Processing document					
32	LogMessage_InitStarted	Project initialization started					
33	LogMessage_DigitizationS...	Digitization started for:					
34	LogMessage_DigitizationF...	Digitization finished					
35	LogMessage_AutoClassific...	Automatic classification started					
36	LogMessage_Classificatio...	Classification finished					
37	LogMessage_AutoClassific...	Automatic classification results:					
38	LogMessage_CreateClass...	Creating classification validation action					
Results							
Messages							
Id	Name	Value	Description	NameAsset	Asset	DescriptionAsset	
39	LogMessage_WaitOrChas...	Waiting on classification validation action #					
40	LogMessage_ShowClassifi...	Presenting Classification Station					
41	LogMessage_ResumeAthe...	Resuming after classification validation [0]. Document(s)...					
42	LogMessage_TrainClassifi...	Train classifiers started					
43	LogMessage_ClassifierTr...	Classification retraining failed					
44	LogMessage_ClassifierTr...	Classifiers trained successfully					
45	LogMessage_DataExtract...	Data extraction started for:					
46	LogMessage_DataExtract...	Data extraction finished					
47	LogMessage_CreateValida...	Creating data validation action for:					
48	LogMessage_WaitOnValid...	Waiting on data validation action #					
49	LogMessage_ShowValidat...	Presenting Validation Station for:					
50	LogMessage_ResumeAthe...	Resuming after data validation [0]. Document: [1]					
51	LogMessage_TrainExtract...	Train extractors started					
52	LogMessage_DataExportS...	Data export started for:					
53	LogMessage_EndProcess...	End process started					
54	LogMessage_ExtractionBu...	Extraction Business Rule Validation started for:					
55	LogMessage_Classificati...	Classification Business Rule Validation started for:					
56	LogMessage_GetWritePer...	Getting file write permission					
57	LogMessage_GiveUpWrit...	Giving up file write permission					

58	LogMessage_LockFile	Locking keyword file:					
59	LogMessage_UnLockFile...	Unlocking keyword file:					
60	LogMessage_UnLockFile...	Unlock: complete!					
61	LogMessage_SetTransact...	Get Transaction Item started					
62	LogMessage_TransactionI...	No transaction item found in the queue					
63	LogMessage_TargetFileM...	Transaction item does not specify the file to be process...					
64	LogMessage_SetTransact...	Transaction Successful					
65	LogMessage_SetTransact...	Transaction failed due to Business Rule Exception:					
66	LogMessage_SetTransact...	Transaction failed due to System Exception:					
67	LogMessage_SetTransact...	Updating transaction progress to:					
68	TransactionProgress_Clas...	Validating Classification					
69	TransactionProgress_Extra...	Validating Extracted Data					
70	WarnMessage_Document...	Failed to process page range (0) in file (1) due to (2) vi...					
71	ErrorMessage_InputDocu...	Cannot find input document:					
72	ErrorMessage_AsyncFailed...	Failed to load asset:					
73	ErrorMessage_ProcessAb...	Process was aborted due to (0) with message: (1) Full ...					
74	LogMessage_InvoicePost...	Starting Invoice Post Processing					
75	LogMessage_InvoicePost...	Parsing the Date has failed!					
76	LogMessage_InvoicePost...	One or more of the mandatory were missing!					

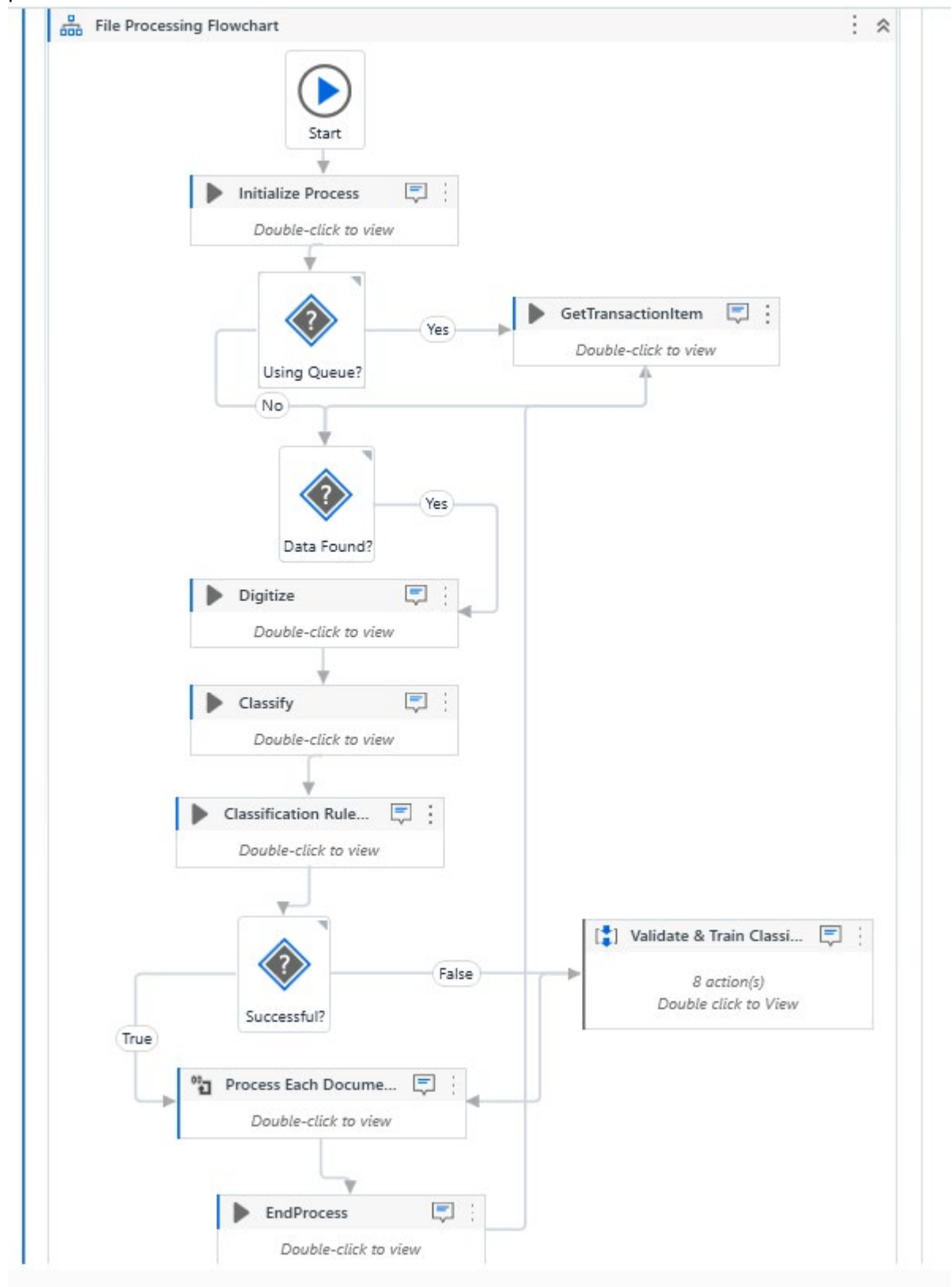
  

Id	Name	Value	Description	NameAsset	Asset	DescriptionAsset
90	Shipping Address-Confiden...	50				
91	Invoice Number-Confidence	85				
92	PO Number-Confidence	85				
93	other-Confidence	50				
94	LogMessage_ReceiptPost...	Starting Receipt Post Processing				
95	LogMessage_ReceiptPost...	Parsing the Date has failed!				
96	LogMessage_ReceiptPost...	One or more of the mandatory were missing!				
97	LogMessage_ReceiptPost...	AutoExtraction is already set to false in a previous step ...				
98	LogMessage_ReceiptPost...	Total Computation Failed!				
99	LogMessage_ReceiptPost...	Subtotal Pass with subtotal: (0)				
100	LogMessage_ReceiptPost...	These are the fields that failed:				
101	1... SubTotalAdditionReceipt	Tax Amount	Each field present in the list here, will be added to t...			
102	1... MandatoryFieldsReceipt	Receipt Number,Tax Amount,Total Value,Receipt Date	These fields are mandatory. If they have not been ...			
103	1... MandatoryColumnFieldsRe...	Line Amount	These column fields are mandatory. If they have no...			
104	1... ConfidenceFieldsReceipt	Receipt Date,Receipt Number	List of fields which will be checked versus a specifi...			
105	1... OtherConfidenceFieldsRec...	Phone Number,Tax Amount,Total Value	List of fields which will be checked versus a single ...			
106	1... Receipt Date-Confidence	50				
107	1... Receipt Number-Confidence	85				
108	1... otherReceipt-Confidence	50				

- Here also If the table isn't created it'll be used a workflow called "Load Config File to SQL" and this will take care of uploading the config to the database in addition to creating the table
- Document Understanding framework works under the concept of a single job, which means that unlike REFramework, it only executes one job or transaction at a time, so you would have to run the performer several times if there are several documents. There are two options, either it is executed several times to process each invoice, or a modification is made to run it once and process all invoices with a single job, this last option is the one we will apply in this case.
- The classification and validation will be done through Action Center.
- It is important to create and edit the ASSETS

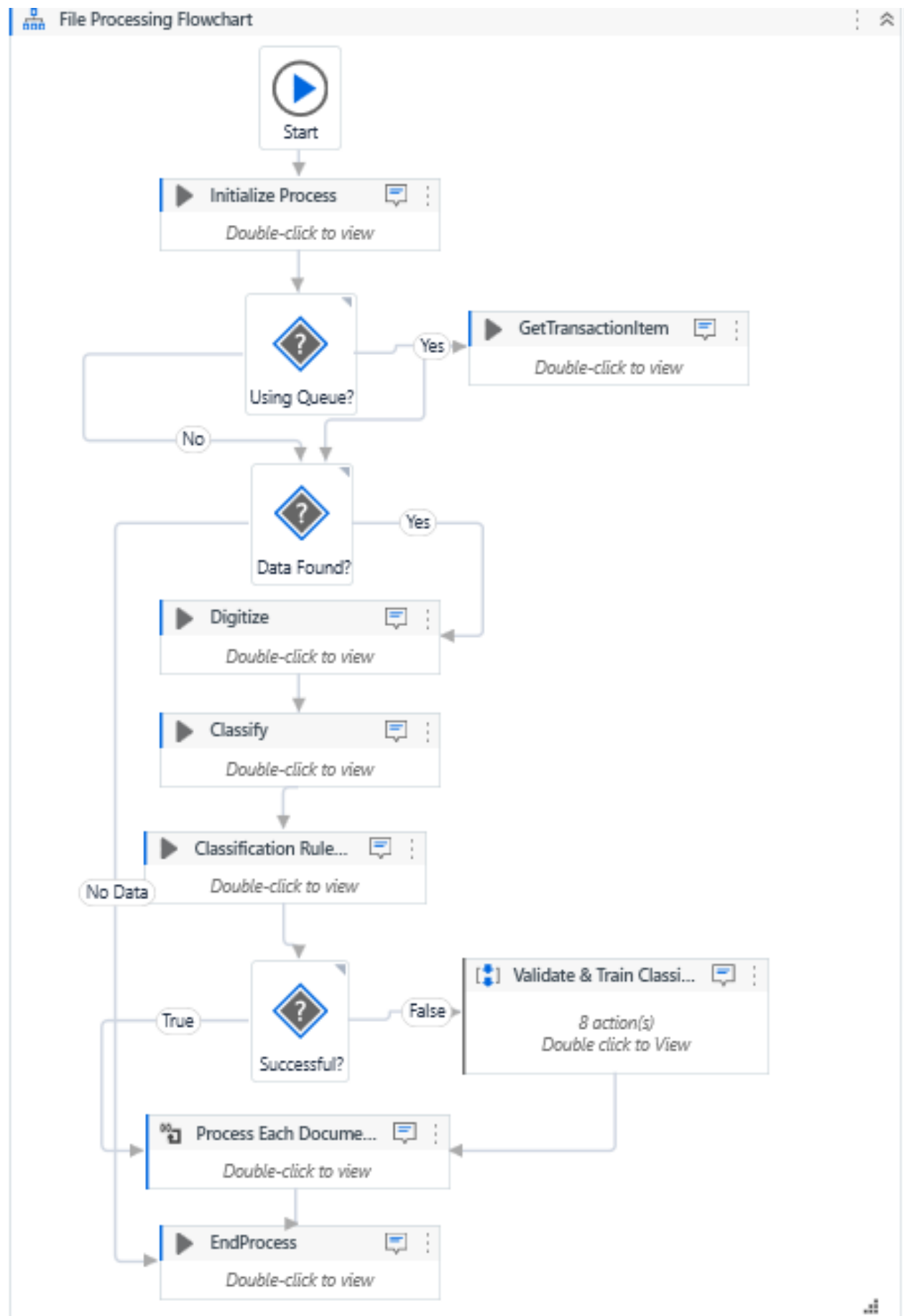
The branch not coming from the Flow decision "Data found?" is eliminated and also the end process branch is connected with the Workflow GetTransactionItem, in this way multiple pdf are

processed.













If the objective is to work under the concept of a "Single job" the workflow should have the following structure:









 DispatcherDocProcessRF_SQL_Bot	11/11/2023 11:27 PM	File folder
 FilesInput	11/12/2023 12:35 AM	File folder
 FilesOutput	11/12/2023 12:51 AM	File folder
 PerformerDocProcessRF_SQL_Bot	11/11/2023 3:49 PM	File folder

- Documents saved from the dispatcher and the trigger

 Certificate of filing.pdf	10/25/2023 9:30 PM	Microsoft Edge P...	93 KB
 Invoice.pdf	10/25/2023 9:30 PM	Microsoft Edge P...	32 KB
 Receipt.pdf	10/25/2023 9:30 PM	Microsoft Edge P...	1,722 KB
 W-9.pdf	10/25/2023 9:30 PM	Microsoft Edge P...	1,209 KB

- When the documents are processed (classified, validated, and exported). It will produce the following exports

 Certificate of filing_1-1.xlsx	10/25/2023 9:32 PM	Microsoft Excel W...	7 KB
 Invoice_1-1.xlsx	10/25/2023 9:34 PM	Microsoft Excel W...	10 KB
 Receipt_1-1.xlsx	10/25/2023 9:37 PM	Microsoft Excel W...	10 KB
 W-9_1-1.xlsx	10/25/2023 9:41 PM	Microsoft Excel W...	8 KB

It is advisable to refer to the video where the complete process is explained step by step.

<https://drive.google.com/file/d/1DSfbOZSYEUUyiYlwsXX0dLP8wT3VoD3f/view?usp=sharing>