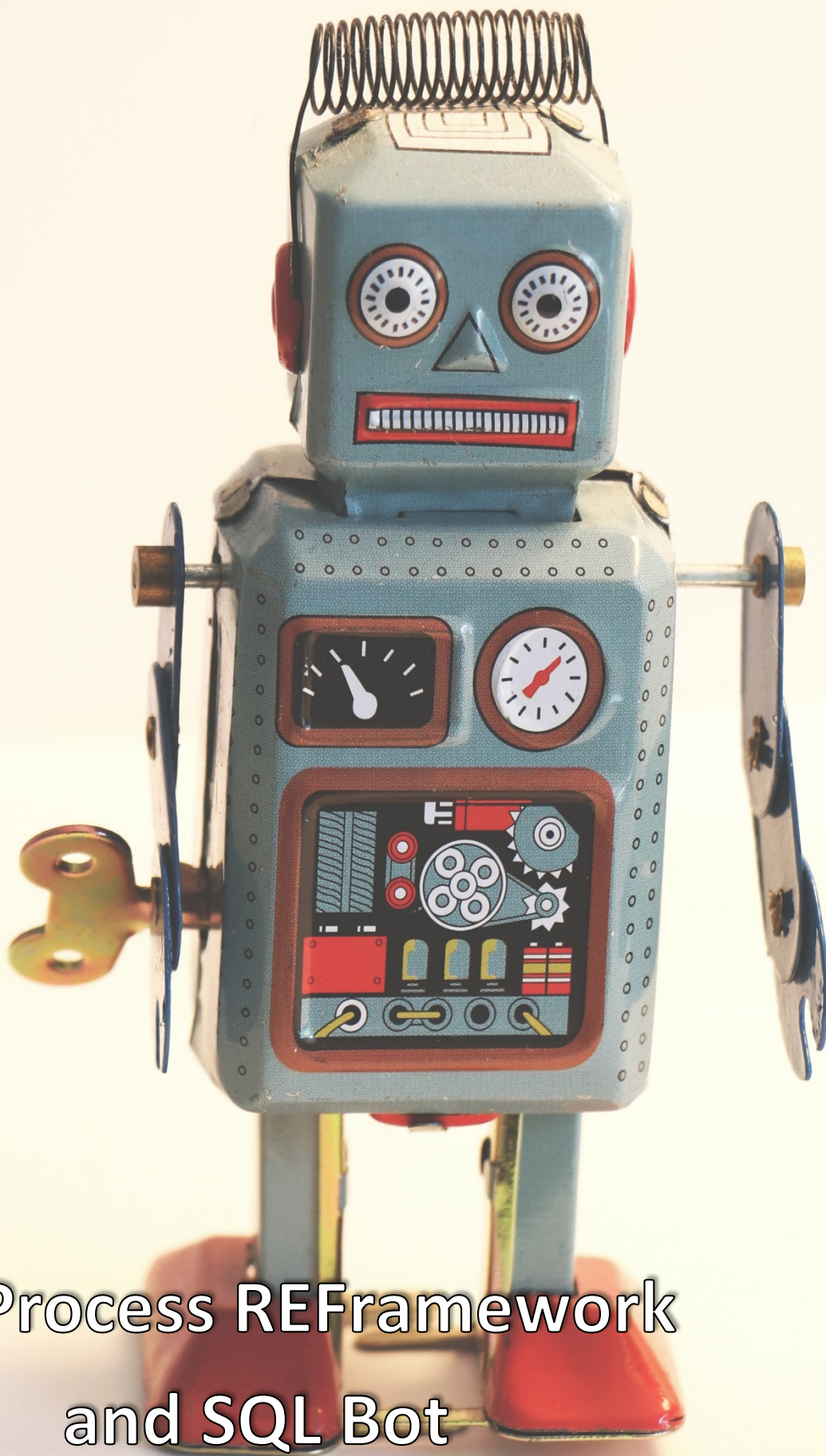


Project by Brayan Jimenez



Doc Process REFramework
and SQL Bot

USE CASE

As is process:

Video de proceso:

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

Una empresa recibe los siguientes tipos de documentos a través de Outlook:

- Invoice
- Receipt
- W9 Form
- Certificate of filing

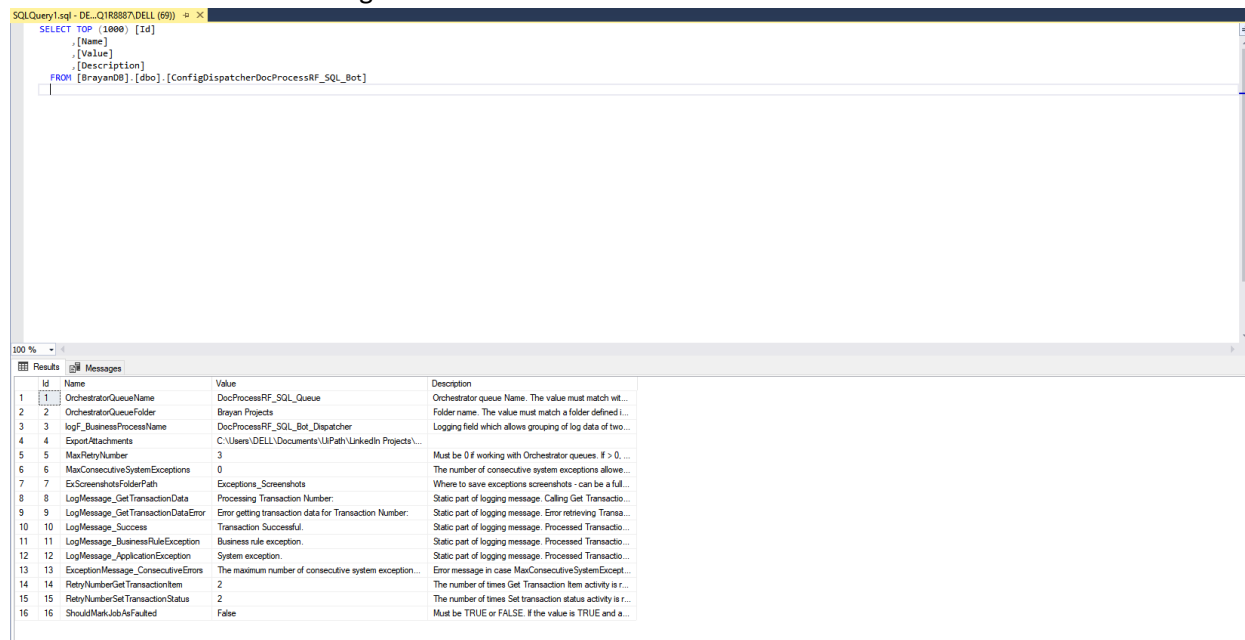
Estos documentos deben ser extraídos del email automáticamente, luego descargados, luego subidos a filas en el orquestador, luego procesados para extraer los datos de cada tipo de documento para su posterior exportación a Excel, la información será extraída de cada archivo pdf y luego exportada en un Excel para cada tipo de documento.

Se deben utilizar triggers, para disparar el dispatcher, y triggers para disparar el performer.

To be process:

Dispatcher:

- Es obligatorio utilizar la base de datos para esta automatización, en este caso para almacenar el archivo de configuración.



```
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... |

- Si la tabla no está creada se utilizará un flujo de trabajo llamado "Load Config File to SQL" que se encargará de cargar la configuración en la base de datos además de crear la tabla.
- Se crea una cola en el orquestador que es donde se cargarán posteriormente todos los documentos.
- The dispatcher se crea mediante REFramework para que el Lineal Process cargue los documentos descargados a través de Outlook en la cola del orquestador.
- Integration service se utiliza para utilizar un activador de eventos a través de una conexión con Outlook 365. Este trigger será utilizado por el dispatcher y permitirá que el

dispatcher sea automáticamente disparado cuando nuevos items, en este caso documentos descargados, sean cargados a la cola. El trigger de eventos supervisa Outlook para recibir un correo electrónico con el asunto "Uipath" y que contenga documentos.

Performer:

- El performer se crea mediante el Document Understanding Framework y se configura los assets, primero se configuran los ajustes y se indican el DocumentUnderstandingQueueName, el Target file, el ExportsFolder, etc. A continuación, se configura la hoja de assets. Para ello, primero se crean todos los assets en el orquestador para poder utilizarlos en la hoja de assets del archivo de configuración.
- Aquí también se utilizará la tabla de base de datos para almacenar el archivo de configuración del performer

SQLQuery2.sql - DE_Q1R8887/DELL (79)

```
SELECT TOP (1000) [Id]
, [Name]
, [Value]
, [Description]
, [NameAsset]
, [Asset]
, [DescriptionAsset]
FROM [BrayanDB].[dbo].[ConfigPerformerDocProcessRF_SQL_Bot]
```

100 %

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\UIPath\Unkined 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 | ActionFolder_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 perform... | | | |
| 14 | SkipExtractorTraining | False | If set to True, extractor training will not be perform... | | | |
| 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... | | | |

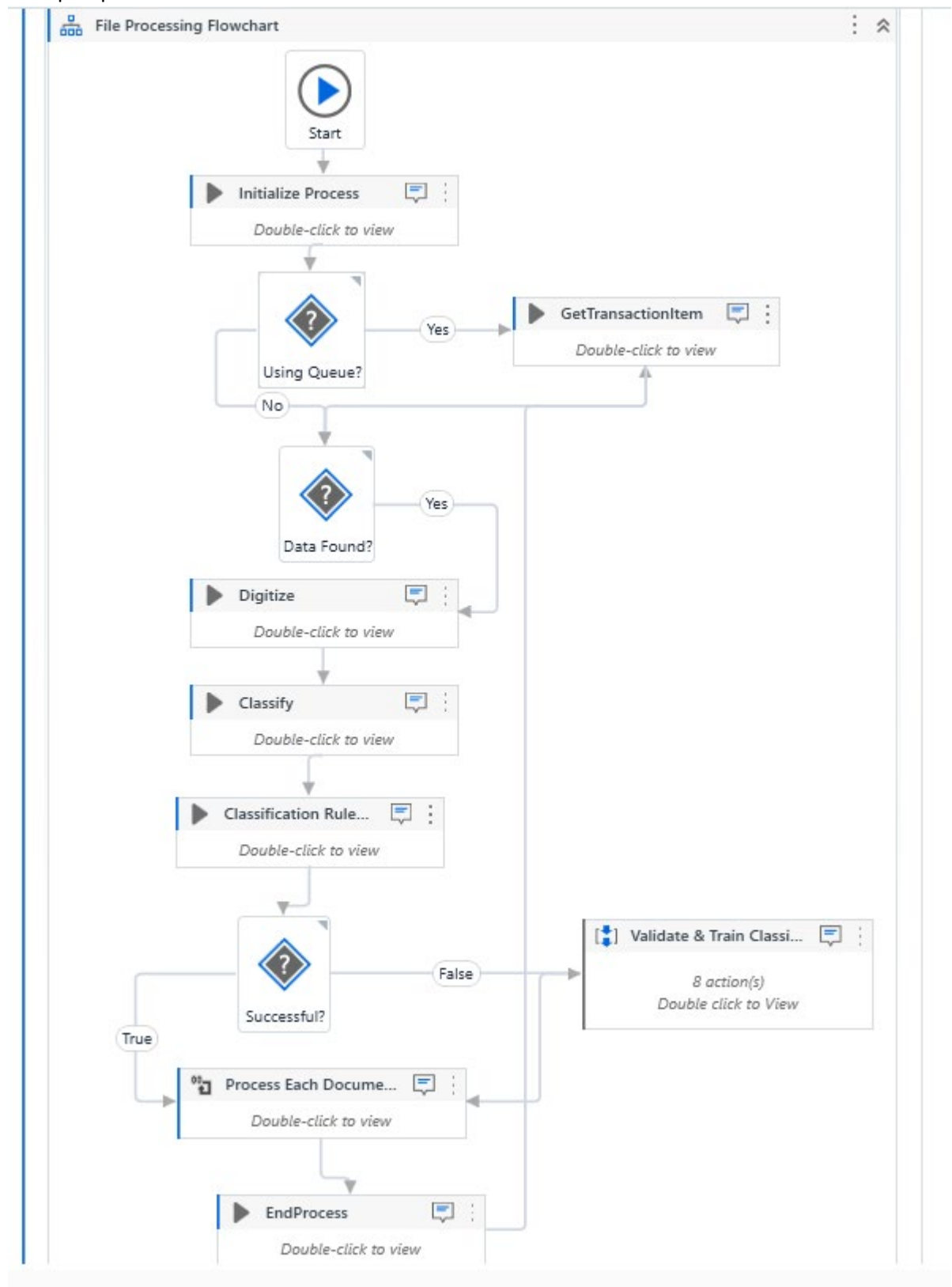
| Id | Name | Value | Description | NameAsset | Asset | DescriptionAsset |
|----|-----------------------------|---|---|-----------|-------|------------------|
| 20 | StorageBucketDirectoryPath | | Path inside the Storage Bucket where actions will ... | | | |
| 21 | DocumentUnderstandingQ... | DU_Invoice | 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 | MaxExecutionAttemptsCa... | 3 | | | | |
| 26 | MaxExecutionAttemptsEtr... | 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_Classificato... | Classification finished | | | | |
| 37 | LogMessage_AutoClassific... | Automatic classification results | | | | |
| 38 | LogMessage_CreateClassi... | Creating classification validation action | | | | |

| Results Messages | | | | | | | |
|------------------|------|------------------------------|---|--|------------|-------|-------------------|
| | Id | Name | Value | Description | Name/Asset | Asset | Description/Asset |
| 39 | 39 | LogMessage_WaitOnClas... | Waiting on classification validation action # | | | | |
| 40 | 40 | LogMessage_ShowClassifi... | Presenting Classification Station | | | | |
| 41 | 41 | LogMessage_ResumeRhe... | Resuming after classification validation(0). Document(s)... | | | | |
| 42 | 42 | LogMessage_TrainClassifi... | Train classifiers started | | | | |
| 43 | 43 | LogMessage_ClassifierTra... | Classification retraining failed | | | | |
| 44 | 44 | LogMessage_ClassifierTra... | Classifiers trained successfully | | | | |
| 45 | 45 | LogMessage_DataExtracti... | Data extraction started for: | | | | |
| 46 | 46 | LogMessage_DataExtracti... | Data extraction finished | | | | |
| 47 | 47 | LogMessage_CreateValida... | Creating data validation action for: | | | | |
| 48 | 48 | LogMessage_WaitOnValid... | Waiting on data validation action # | | | | |
| 49 | 49 | LogMessage_ShowValidat... | Presenting Validation Station for: | | | | |
| 50 | 50 | LogMessage_ResumeRhe... | Resuming after data validation(0). Document: (1) | | | | |
| 51 | 51 | LogMessage_TrainExtract... | Train extractors started | | | | |
| 52 | 52 | LogMessage_DataExportS... | Data export started for: | | | | |
| 53 | 53 | LogMessage_EndProcess... | End process started | | | | |
| 54 | 54 | LogMessage_ExtractionBu... | Extraction Business Rule Validation started for: | | | | |
| 55 | 55 | LogMessage_Classificati... | Classification Business Rule Validation started for: | | | | |
| 56 | 56 | LogMessage_GetWritePer... | Getting file write permission | | | | |
| 57 | 57 | LogMessage_GiveUpWrit... | Giving up file write permission | | | | |
| 58 | 58 | LogMessage_LockFile... | Locking keyword file: | | | | |
| 59 | 59 | LogMessage_UnlockFile... | Unlocking keyword file: | | | | |
| 60 | 60 | LogMessage_UnlockFile... | Unlock complete! | | | | |
| 61 | 61 | LogMessage_GetTransact... | Get Transaction Item started | | | | |
| 62 | 62 | LogMessage_TransactionI... | No transaction item found in the queue | | | | |
| 63 | 63 | LogMessage_TargetFileM... | Transaction item does not specify the file to be process... | | | | |
| 64 | 64 | LogMessage_SetTransact... | Transaction Successful | | | | |
| 65 | 65 | LogMessage_SetTransact... | Transaction failed due to Business Rule Exception: | | | | |
| 66 | 66 | LogMessage_SetTransact... | Transaction failed due to System Exception: | | | | |
| 67 | 67 | LogMessage_SetTransact... | Updating transaction progress to: | | | | |
| 68 | 68 | TransactionProgress_Clas... | Validating Classification | | | | |
| 69 | 69 | TransactionProgress_Extra... | Validating Extracted Data | | | | |
| 70 | 70 | WarnMessage_Document... | Failed to process page range (0) in file (1) due to (2) vi... | | | | |
| 71 | 71 | ErrorMessage_InputDocu... | Cannot find input document: | | | | |
| 72 | 72 | ErrorMessage_AsyncFai... | Failed to load asset: | | | | |
| 73 | 73 | ErrorMessage_ProcessAb... | Process was aborted due to (0) with message: (1) Full ... | | | | |
| 74 | 74 | LogMessage_InvoicePost... | Starting Invoice Post Processing | | | | |
| 75 | 75 | LogMessage_InvoicePost... | Parsing the Date has failed! | | | | |
| 76 | 76 | LogMessage_InvoicePost... | One or more of the mandatory were missing! | | | | |
| 90 | 90 | Shipping Address-Confiden... | 50 | | | | |
| 91 | 91 | Invoice Number-Confidence | 85 | | | | |
| 92 | 92 | PO Number-Confidence | 85 | | | | |
| 93 | 93 | other-Confidence | 50 | | | | |
| 94 | 94 | LogMessage_ReceiptPost... | Starting Receipt Post Processing | | | | |
| 95 | 95 | LogMessage_ReceiptPost... | Parsing the Date has failed! | | | | |
| 96 | 96 | LogMessage_ReceiptPost... | One or more of the mandatory were missing! | | | | |
| 97 | 97 | LogMessage_ReceiptPost... | AutoExtraction is already set to false in a previous step ... | | | | |
| 98 | 98 | LogMessage_ReceiptPost... | Total Computation Failed! | | | | |
| 99 | 99 | LogMessage_ReceiptPost... | Subtotal Pass with subtotal: (0) | | | | |
| 100 | 1... | LogMessage_ReceiptPost... | These are the fields that failed: | | | | |
| 101 | 1... | SubTotalFieldsReceipt | 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 | | | | |

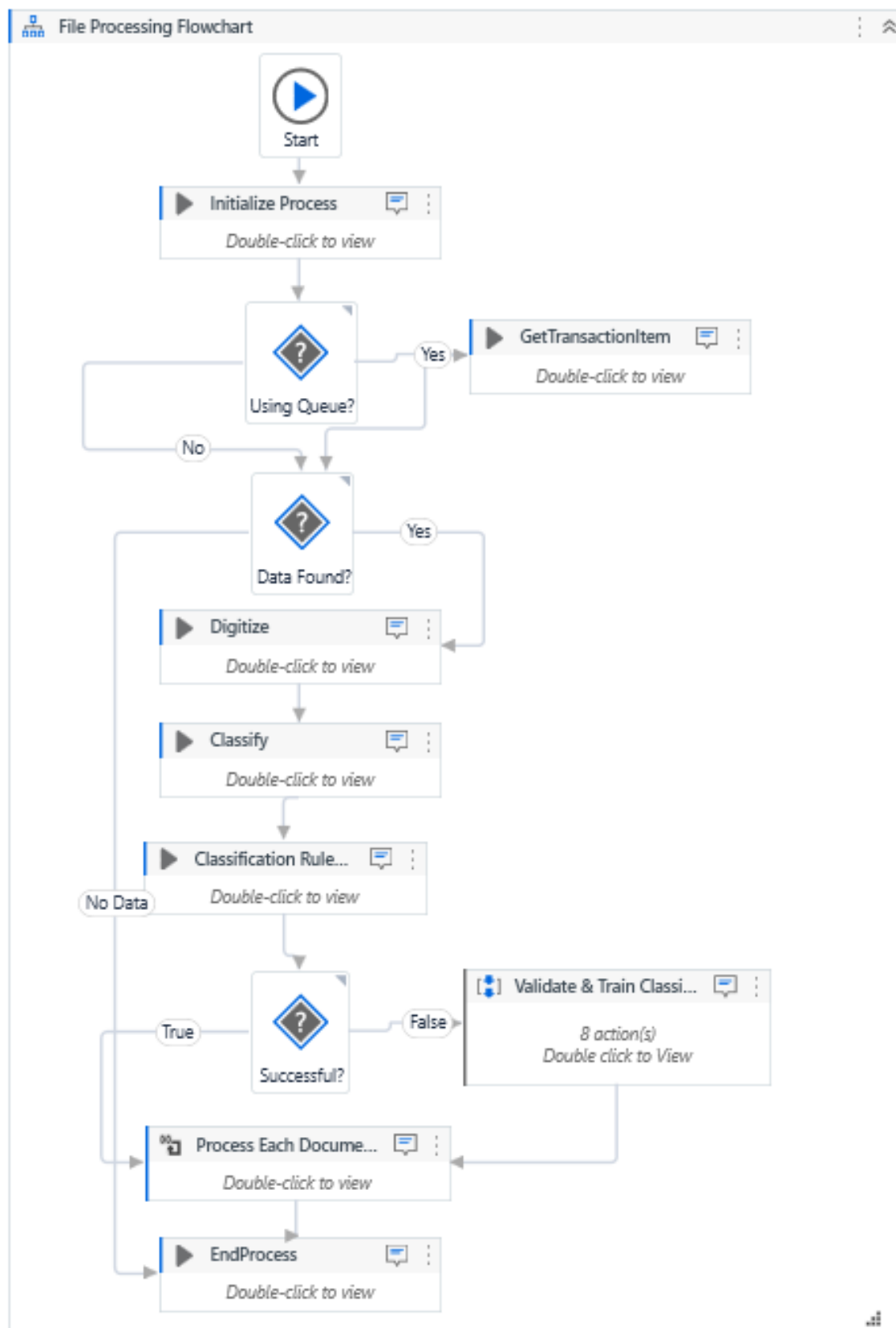
-
-
- Aqui Tambien si la tabla no esta creada será usado un workflow llamado “Load config file to SQL” y esto se encargará de subir el config a la base de datos además de crear la tabla.
- Document Understanding framework trabaja bajo el concepto de “a single job”, lo que significa que a diferencia del REFramework, solo ejecuta un job o transacción a la vez, asi que tendrías que correr el performer varias veces si hay varios documentos. Hay dos opciones, ya sea es ejecutado varias veces para procesar cada invoice, o una modificación es hecha para correrlo una sola vez y procesar todas las invoices con un solo job, esta ultima opción es la única que aplicaremos en este caso.
- La clasificacion y validacion será hecho a traves de action center.
- Es importante crear y editar los assets.





Se elimina la rama que no proviene de la decisión del Flujo "¿Data found?" y también se conecta la rama del proceso final con el Flujo de Trabajo GetTransactionItem, de esta forma se procesan

múltiples pdf.







Si el objetivo es trabajar bajo el concepto de “a single job ”el workflow debería seguir la siguiente estructura (la cual es por defecto):







| | | |
|--|---------------------|-------------|
|  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 |

- Documentos guardados obtenidos por medio de Outlook con ayuda del trigger y dispatcher

| | | | |
|---|--------------------|---------------------|----------|
|  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 |

- Cuando los documentos son procesados (clasificados, validados, y exportados. Se producirá las siguientes extracciones.

| | | | |
|--|--------------------|----------------------|-------|
|  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 |

Es recomendable dirigirse al video donde el proceso completo es explicado paso por paso.

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