

Automation Document Processing Lab

## **IBM Cloud Pak for Business Automation Demos and Labs 2022**

### **Capture**

IBM Automation Document Processing  
V22.0.2

### **Lab Automation Document Processing**

V 2.0

Clandis Baker  
SWAT Business Automation Portfolio Specialist – Capture Products  
[bakercl@us.ibm.com](mailto:bakercl@us.ibm.com)

Krish Lakshminarayanan  
Global Technical Program Leader for Capture / Intelligent Document Processing Global Sales (WW)  
[krishkrish@ibm.com](mailto:krishkrish@ibm.com)

Ryan Sparks  
Advisory Business Automation Tech Sales Leader – RPA/ADP  
[rmsparks@us.ibm.com](mailto:rmsparks@us.ibm.com)

## NOTICES

This information was developed for products and services offered in the USA.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119

Armonk, NY 10504-1785

United States of America

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk. IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements, or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

## TRADEMARKS

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is

available on the web at "Copyright and trademark information" at  
[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a Registered Trade Mark of AXELOS Limited.

ITIL is a Registered Trade Mark of AXELOS Limited.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

© Copyright International Business Machines Corporation 2020.

This document may not be reproduced in whole or in part without the prior written permission of IBM.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Table of Contents

<b>1. Overview.....</b>	<b>6</b>
1.1 Getting HELP during the lab.....	6
1.2 Icons .....	6
1.3 Abstract .....	6
1.4 Introduction.....	7
<b>2 Getting started .....</b>	<b>8</b>
2.1 IBM TechZone – Overview.....	8
2.1.1 Reserve Environment .....	9
2.2 Open your IBM Cloud Environment .....	12
<b>3 Lab Overview.....</b>	<b>14</b>
3.1 How does ADP work? .....	14
<b>4 Create Document Processing Project .....</b>	<b>16</b>
4.1 Reviewing the interface.....	21
4.1.1 Build Tab .....	21
4.1.2 Enrich Tab .....	22
4.1.3 Configure Tab.....	23
<b>5 Configure a Wage and Tax document type. ....</b>	<b>26</b>
5.1 Create Wage and Tax document type.....	26
5.2 Create Field .....	28
5.3 Create the Employee Name Address field.....	33
5.4 Create Employee Social Security Number Field.....	34
<b>6 Document Types and Samples Overview.....</b>	<b>37</b>
6.1 Categorize documents. ....	37
<b>7 Train classification.....</b>	<b>45</b>
7.1 How do I improve my results? .....	50
7.1.1 Option 1 – Add more samples.....	50
7.1.2 Option 2 – Review all uploaded samples.....	51
<b>8 Data extraction.....</b>	<b>52</b>
8.1 Correcting extracted values.....	56
8.2 Train extraction model .....	61
<b>9 Data standardization .....</b>	<b>62</b>
<b>10 Version and deploy your project .....</b>	<b>64</b>
<b>11 Application designer .....</b>	<b>67</b>
11.1 Create your Runtime Application .....	67
11.2 Upload documents for processing .....	74
11.3 Correct any classification errors.....	77
11.4 Correct extraction issues.....	79
<b>12 Optional Export/Import Project.....</b>	<b>85</b>
<b>Appendix A - Troubleshooting .....</b>	<b>87</b>
TechZone Pending Status taking Long Time .....	87
Service Error .....	87
Application Blank .....	88

Connection issue with Workstation to Cloud.....	89
Opening an Incognito Window.....	89
Popup Blocked when trying to Preview Application. ....	90
<b>Appendix B - BAW &amp; ADP Integration Sample .....</b>	<b>91</b>
<b>Appendix C - Badge Information. ....</b>	<b>92</b>

## 1. Overview

### 1.1 Getting HELP during the lab

- For internal IBM, another good resource is the Archive slack channel for questions: #cp4ba-adp-lab or <https://ibm-cloud.slack.com/archives/C01LVVBMWPN>
- For external participants besides the Slack channel, use the Webex chat if you are in a webex event or just speak up.
- For others, email bakercl@us.ibm.com. This method will be slower and will be best effort. It may require jumping on a Webex meeting to provide help.
- Getting help after lab reach out to the following:
  - [bakercl@us.ibm.com](mailto:bakercl@us.ibm.com)
  - [krishkirsh@us.ibm.com](mailto:krishkirsh@us.ibm.com)
  - [rmsparks@us.ibm.com](mailto:rmsparks@us.ibm.com)

### 1.2 Icons

The following symbols appear in this document at places where additional guidance is available.

Icon	Purpose	Explanation
	Important!	This symbol calls attention to a particular step or command. For example, it might alert you to type a command carefully because it is case sensitive.
	Information	This symbol indicates information that might not be necessary to complete a step but is helpful or good to know.
	Trouble-shooting	This symbol indicates that you can fix a specific problem by completing the associated troubleshooting information.

•

### 1.3 Abstract

Set up a capture solution in minutes. Introduce technical sellers to IBM Automation Document Processing. In this session, students will configure their own capture project. They will learn how to use machine learning classification for their sample documents, define fields for extraction, create validation rules, and use deep learning\* (subject to environment configuration) to automate data extraction.

## 1.4 Introduction

Welcome to the Automation Document Processing lab. This lab will introduce you to Document Processing and provide you with an understanding how you can configure it for your customer opportunities.

Automation Document Processing provides a tailored solution that reads your documents (in English, French, Spanish, German, Dutch, Portuguese), extracts data, and refines and stores the data for use.

With the right business knowledge, you can design deep learning models without being a data scientist. The Document Processing Designer includes pre-trained deep learning models that you can use as a base for your own model. The pre-trained document types include bills of lading, invoices, and utility bills.

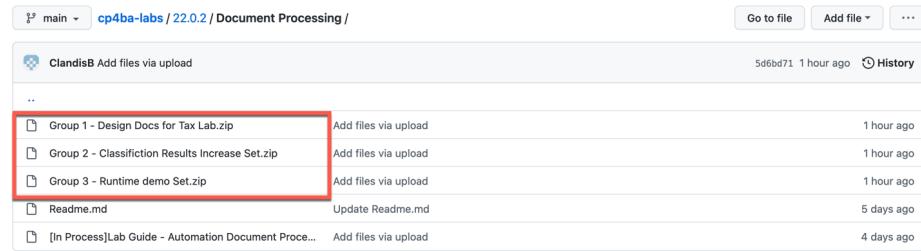
You can extract text, check boxes, forms, tables, barcodes, signature detection and even free text. With no or low code options, you can create an application that processes documents, extracts data, flags issues, and stores your documents and data. And the data enrichment capabilities ensure that the extracted data is standardized and ready for use in downstream integrations.

This lab will not cover all the available functionality available due to time constraints. Additional labs will be created in the next few months to add to your knowledge and understanding of Document Processing.

## 2 Getting started

Download the sample documents in the zip file. We will be using these sample documents during the labs You can find them here:

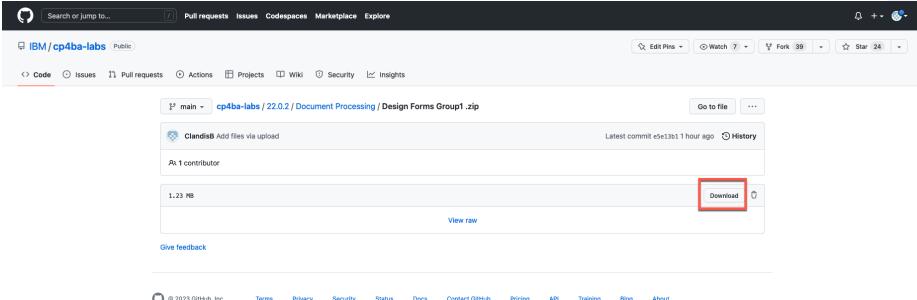
<https://github.com/IBM/cp4ba-labs/tree/main/22.0.2/Document%20Processing>



File	Action	Last Modified
Group 1 - Design Docs for Tax Lab.zip	Add files via upload	1 hour ago
Group 2 - Classification Results Increase Set.zip	Add files via upload	1 hour ago
Group 3 - Runtime demo Set.zip	Add files via upload	1 hour ago
Readme.md	Update Readme.md	5 days ago
[In Progress]Lab Guide - Automation Document Proce...	Add files via upload	4 days ago

\_1. Click on “Group1 – Design Docs for Tax Lab.zip”.

\_2. Then Click on Download



\_3. Repeat above steps “Group 2 – Classification Results Increase Set.zip” and “Group 3 – Runtime Set.zip”

\_4. Unzip the files and keep them in their designated folder.

You will notice the images are in various unique folders that will be referenced specifically in the different labs later. Please keep them in their proper folders.

### 2.1 IBM TechZone – Overview

What is IBM TechZone?

IBM Technology Zone ([techzone.ibm.com](http://techzone.ibm.com)) enables IBM teams and IBM Business Partners to provision technical “Show Me” live environments, Proof-of-Technologies, prototypes, and Minimum Viable Prototypes, which can be customized, shared with peers and clients to experience IBM Technology.

## Automation Document Processing Lab

Learn more: <https://techzone.ibm.com/collection/onboarding#tab-1>

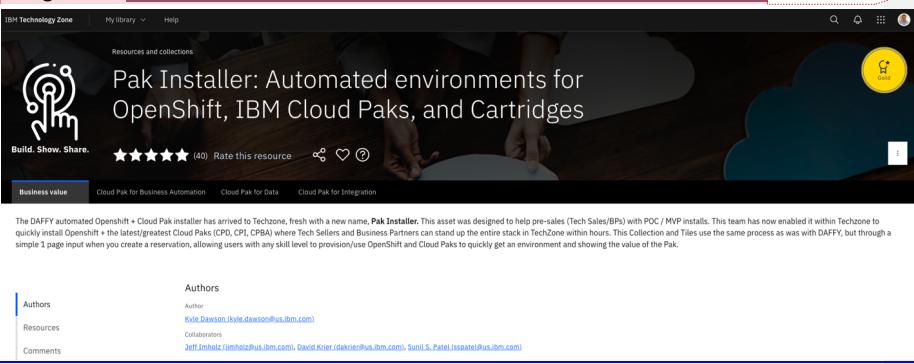
The TechZone leverages DAFFY. DAFFY is **Deployment Automation Framework For You**. The DAFFY installer tool has been renamed to Pak Installer. This tool will do all the heavy lifting of the OpenShift and IBM Cloud Pak installs. The National Market Top Team created Pak Installer to assist the technical sales teams with the progression of IBM Cloud Pak opportunities.

The goal is to provide the technical sales with a set of (easy to use) scripts that will aid in the installation of OpenShift and the IBM Cloud Pak's. For more information on DAFFY/Pak Installer please look at: <https://ibm.github.io/daffy/Deploying-OCP/TechZoneTilesBeta/>

Also please rate the tile on the TechZone website. This will help the development team understand its value.

### 2.1.1 Reserve Environment

\_1. Navigate to <https://techzone.ibm.com/collection/63457fcba311ed0018ca2442> Commented [CB1]: New link needed



The screenshot shows a collection page on the TechZone website. The title is "Pak Installer: Automated environments for OpenShift, IBM Cloud Paks, and Cartridges". It features a 5-star rating and a note about the Daffy automated OpenShift + Cloud Pak installer. Below the main content, there are sections for Authors, Resources, and Comments.

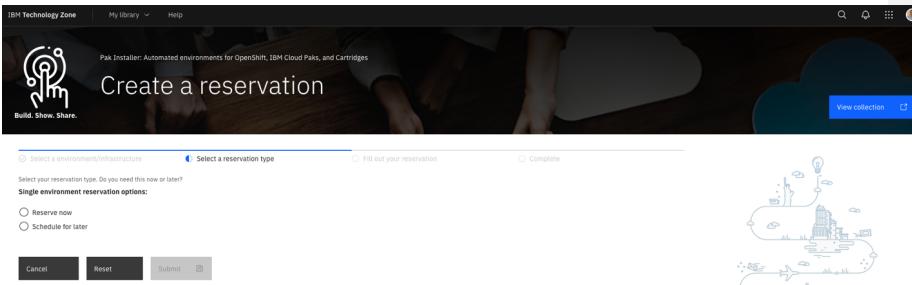


*Note: Don't grab just any CP4BA in TechZone. This environment is built with Daffy authored by Kyle Dawson with the latest releases. This environment can also be used at a customer site with same tool and framework of Daffy.*

- \_2. Click Cloud Pak for Business Automation tab and scroll down to the "Cloud Pak for Business Automation 22.0.2 – VMWare tile.
- \_3. Click on Reserve

## Automation Document Processing Lab

- \_4. On Create a reservation screen **select option** for when to start



- \_5. Create a Reservation

Based on the reservation type you are making, provide the required information

*Customer Demo* : Need a short customer-facing demonstration

*Practice/Self-Education*: Need to gain experience

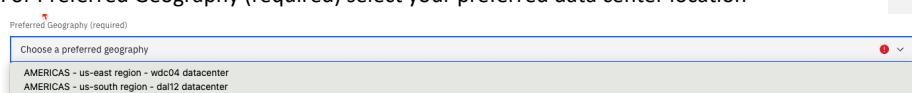
*Standard proof of concept*; Need an environment for a standard product use case.

*Custom Proof of concept*: Need a complex, customized environment.

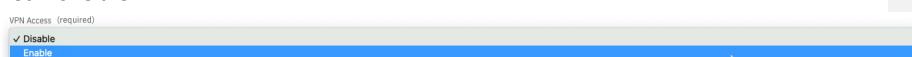
*Testing*: Need to test a specific function, configuration, or customization.

- \_6. For this lab **Select Testing** will give you 3 days plus the option to extend it for another week. Otherwise, you will need a legitimate opportunity to leverage another reservation type.

- \_7. For Preferred Geography (required) select your preferred data center location



- \_8. For VPN Access **choose Enable**. You will be using a VPN to connect from desktop to the TechZone tile

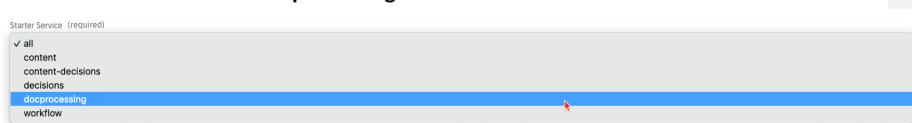


*Make sure to pick enable otherwise you'll have to start all over with deployment.*

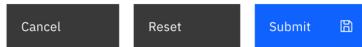
- \_9. In Cloud Pak for Business Automation Version Pick 22.0.2. (if not already chosen for you)

- \_10. In Cloud Pak for business Automation IFix pick IF002 (if not already chosen for you)

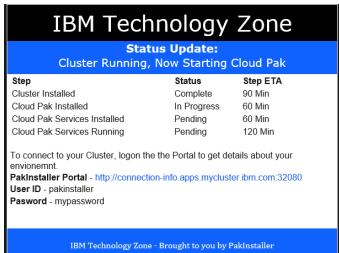
- \_11. For Starter Service **choose docprocessing**



\_12. Click Submit



You will receive emails on the status and the steps that have completed and please allow up to 1 hour for the start-up services to fully complete. Your emails will look like the following:



Once "Cloud Pak Services Running" is Complete upon receiving this email then your environment is ready. Once the start-up process is complete you can click on the links identified in the email. However, it is recommended that you review your reservation information from the IBM Technology Zone – My reservation site.

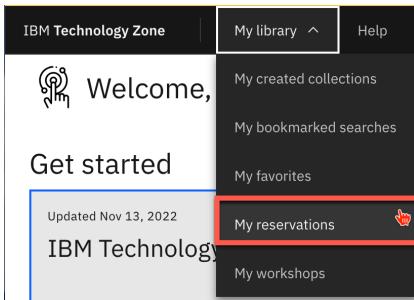


If after receiving email and a few hours have passed and your environment is not up, check [Appendix A – Troubleshooting](#) for possible fix.

\_13. Click My reservations



\_14. Once you get the email from the IBM Technology Zone site, you can access your environment reservation(s) by clicking on the **My library** then **My Reservations**

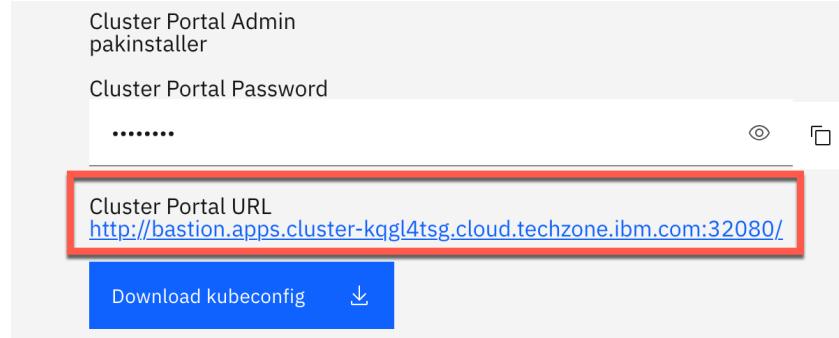


## Automation Document Processing Lab

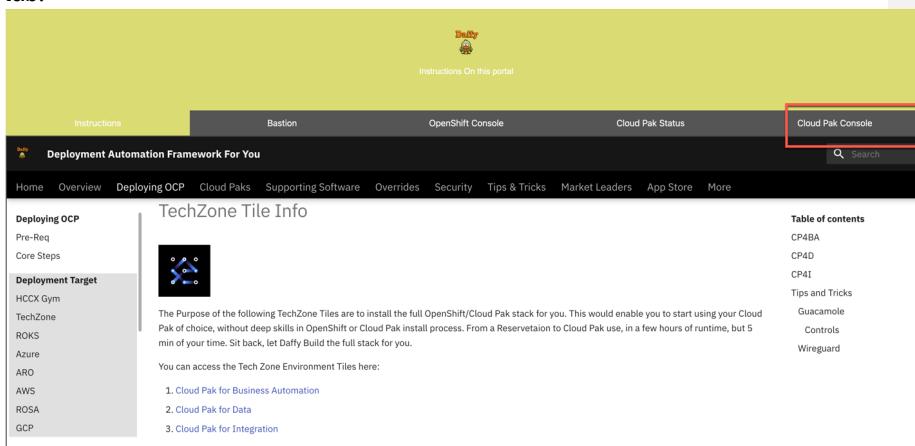
You can also access directly using the link below  
<https://techzone.ibm.com/my/reservations>

### 2.2 Open your IBM Cloud Environment

- \_1. Back on your reservation screen under the Environment section **Right Click** on the link under **Cluster Portal URL** and **Select open in a private window**

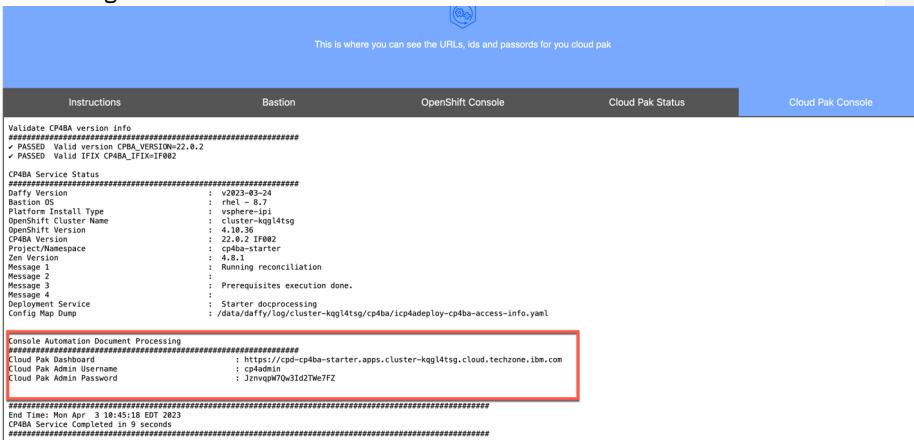


- \_2. Login with <**Cluster Portal Admin /Cloud Pak Admin Password**> provided in Environment section above. In example above pakinstaller/<your unique password>.
- \_3. You will be presented with the Daffy portal screen. Click on **Cloud Pak Console** tab.



## Automation Document Processing Lab

- \_4. Under the Cloud Pak Console tab scroll down to Console Automation Document Processing

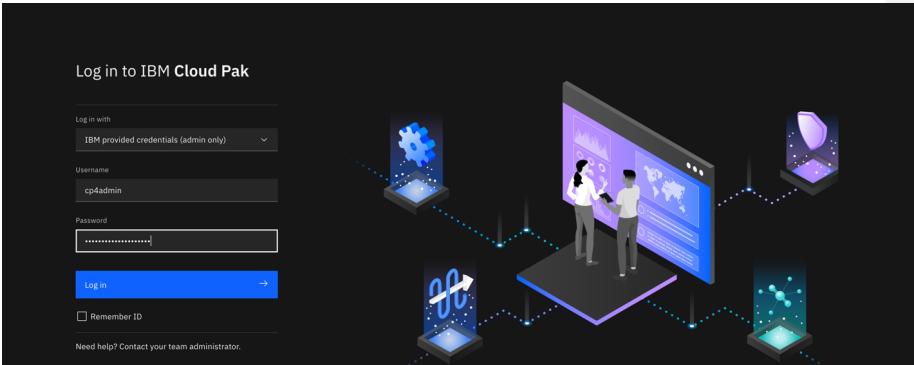


This is where you can see the URLs, ids and passwords for your cloud pak

Instructions	Bastion	OpenShift Console	Cloud Pak Status	Cloud Pak Console
Validate CP4BA version <version>				
CP4BA Service Status				
Cloud Pak Version	: 22.0.2.1F002			
Bastion OS	: rhel - 8.7			
Platform Install Type	: vsphere-ipi			
OpenShift Cluster Name	: cluster-kqgl4tsg			
OpenShift Version	: 4.10.36			
CP4BA Version	: 22.0.2.1F002			
Prerequisites	: cp4ba-starter			
Zen Version	: 4.8.3			
Message 1	: Running reconciliation			
Message 2	: Prerequisites execution done.			
Message 3				
Message 4				
Deployed Services				
Config Map Dump	: /data/daffy/log/cluster-kqgl4tsg/cp4ba/icp4adeploy-cp4ba-access-info.yaml			
Console Automation Document Processing				
Cloud Pak Dashboard	: https://cpd-cp4ba-starter.apps.cluster-kqgl4tsg.cloud.techzone.ibm.com			
Cloud Pak Admin Username	: cp4admin			
Cloud Pak Admin Password	: JzmvqM70x3Id2TMe7FZ			

End Time: Mon Apr 3 18:45:18 EDT 2023  
CP4BA Service Completed in 9 seconds

**Copy** the link to your favorite *private browser* or open in new tab. Accept any potential security risks. You will then be prompted with Log in screen to the IBM Cloud Pak.



- \_5. Enter the **<Cloud Pak Admin Username>** and **<Cloud Pak Admin Password>** from Cloud Pak Console tab. **Click** on blue **Login** button.
- \_6. You will be presented with the “Welcome! Let’s get started” screen. **Click** Maybe later button.



Note you will see this screen several times throughout the lab. You can always select Maybe later while doing this lab.

### 3 Lab Overview

The lab will focus on the design time tasks for Automation Document Processing (ADP). Despite the push for the digitization of content for many years, there are still a lot of paper documents that require workers to read and interpret the information – whether it is structured data, such as tax forms, or semi-structured data, such as invoices, utility bills, and so on. This lab describes how to set up an automate document processing pipeline using ADP.

#### 3.1 How does ADP work?

##### Document Processing Designer

You use the Designer interface to create a set of document types and related fields that comprise your Document Processing project. Document Processing Designer combines an intuitive interface with a set of AI and deep learning tools that identify and learn the document types that matter to your organization. For each document type, you designate which pieces of information to extract as data for that document to be used by downstream applications. You can also apply tools to clean up and standardize the data as it is extracted.

##### Deployment tools

After you build the Document Processing project in the Designer, you deploy the project to make it available for building your document processing application. The deployment process is also used to configure the repository to receive the processed documents from your end-user application by making the capabilities and artifacts available for integration into an application and into the destination repository.

##### Application templates and toolkits

You use the no- or low-code application building capabilities of Application Designer, customized templates and toolkits, and the AI model of your Document Processing project to create a document processing end-user application. This application recognizes your documents, extracts your relevant data, and presents issues to fix before sending the documents to storage and using the data in other systems.

##### Document processing application and document management

The application that you build uses the AI and deep learning to automatically detect, extract, and standardize the data in all your documents. Any anomalies are flagged according to your customized model and the priority that you set so that your document processing user can correct issues before the documents are finalized.

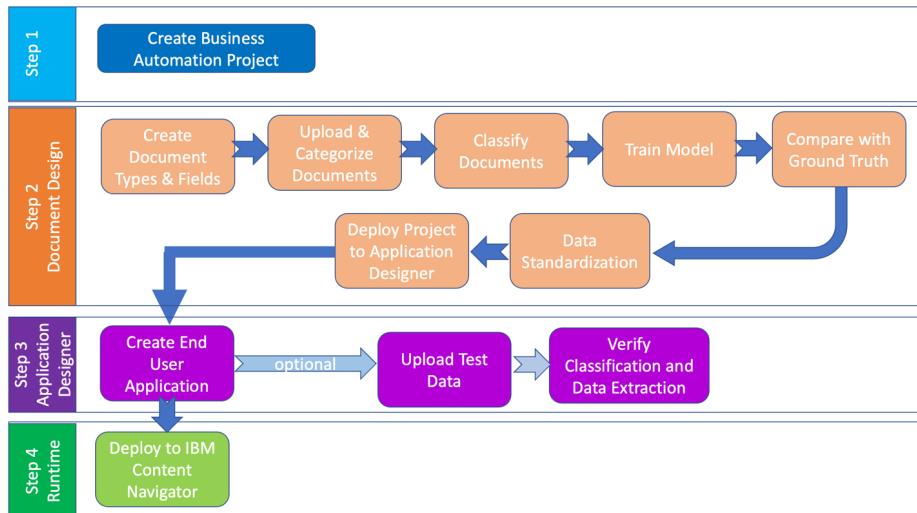
When you deploy your document processing application, you connect it to a content repository that manages the document types and the extracted data for each document. The solution is fully integrated with IBM FileNet® Content Manager, simplifying document and data storage by applying your existing filing architecture and business rules to each processed document. The content and metadata are automatically saved in FileNet within the appropriate document class.

##### End result

## Automation Document Processing Lab

Your document types are stored in the content repository, with appropriate retention and access controls. An associated JSON file reflects all the extracted data for the document. Properties are set on the document with the data definition-controlled values. Your extracted data is cleaned, standardized, and ready for use in other applications.

The following diagram shows the tasks required to configure and deploy a new ADP project.



### Step 1 – Create an ADP Business Automation Project

Each document processing project requires a separate repository in your Git organization. Coordinate with your Git administrator to create the repository for your project.

### Step 2 – Document Design

This step shows the high-level tasks that will be needed to complete to train the system to recognize document types, successfully extract fields and tables, configure the fields in FileNet and finally deploying your content project to the application designer so you can configure the end-user interfaces.

### Step 3 – Application Designer

The application designer is where you would configure end-user interfaces such as the classification and verification screens. The lab will not go in a lot of details on how to configure the interfaces. It will instead show you how to create an application, and test processing a batch of documents through the system. To get more information on creating/using the Business Automation Application (BAA) look at the SWAT Jam Lab for BAA.

### Step 4 – Runtime

End-users would be using the runtime IBM Content Navigator interface to process documents or batches, classify document and verify extracted field data in the verification screen.

## 4 Create Document Processing Project

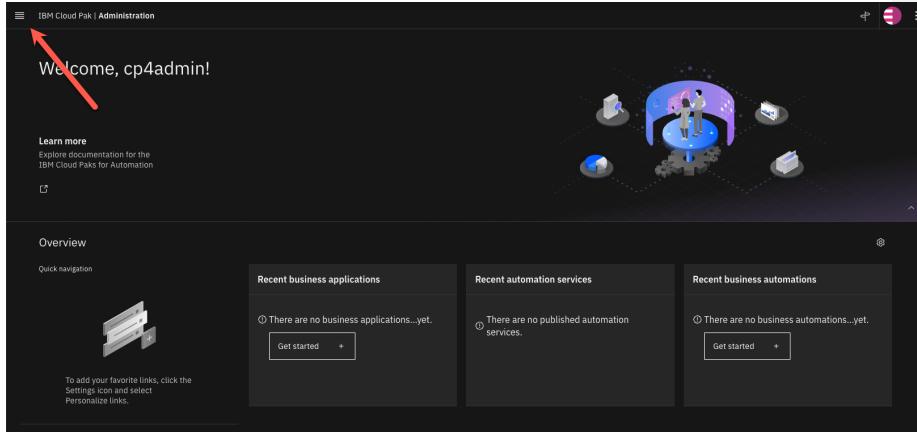
Step 1

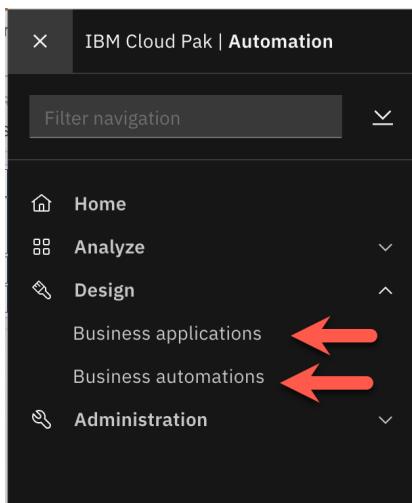
Create Business Automation Project

Cloud Pak for Business Automation Studio is the single authoring and development environment for the IBM Cloud Pak for Business Automation platform that accelerates digital transformation. Business Automation Studio provides an entry point to various designers to help you reach your goals.

There are two distinct parts to the Business Automation Studio configuration.

- \_1. Click on the hamburger menu at the top left next to IBM Automation.





**Business automations** provides the Document Processing configuration of the document classes, and the **Business applications** provides the user interfaces.

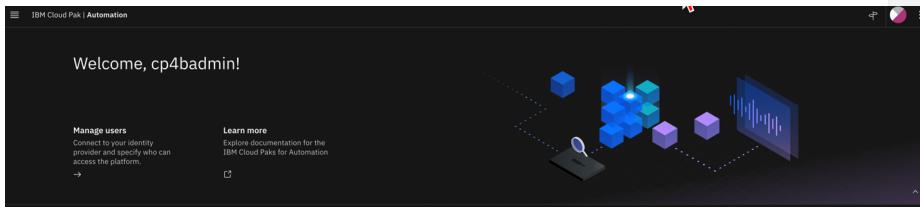
Within the Business Automations you can create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way. Also in Business Automation, you use the **Document Designer** interface within Automations to create a set of document types and related fields that comprise your Document Processing project.

The Document Processing Designer combines an intuitive interface with a set of AI and deep learning tools that identify and learn the document types that matter to an organization. For each document type, you designate which pieces of information to extract as data for that document to be used by downstream applications. You can also apply tools to clean up and standardize the data as it is extracted.

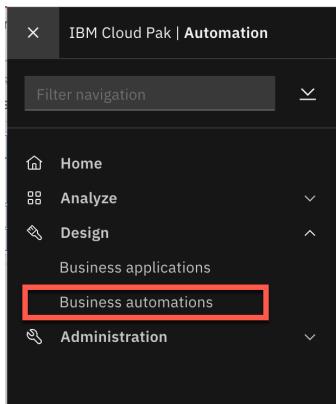
Within **Business Applications** you can quickly create user interfaces that integrate tasks, data, and automations. You can start with a template to ensure consistency. You can also use toolkits to share artifacts from existing applications.

We will start with the Business Automations.  
Once logged in to the IBM Automation Server, you should see the Welcome screen.

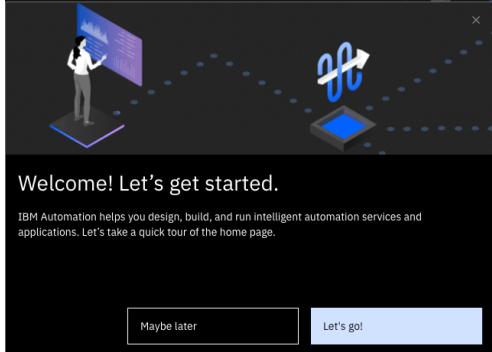
## Automation Document Processing Lab



- \_2. Click on Drop down arrow next to Design then **Select Business Automations**.



You may be presented with an overview screen. Select **Maybe Later**.



## Automation Document Processing Lab

Then following screen appears.

The screenshot shows the 'Business automations' page in the IBM Cloud Pak | Automation interface. At the top, there is a navigation bar with the title 'IBM Cloud Pak | Automation'. Below the title, the page header reads 'Business automations' and includes a brief description: 'Create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way.' A 'Learn more' link is also present. Below the header, there is a toolbar with buttons for 'Create' (highlighted in blue), 'Import', and a download icon. A sidebar on the left lists categories: 'Published automation services' (with an arrow icon), 'Decision', 'Document processing', 'Workflow', and 'External'. Each category has a corresponding arrow icon to its right.

- \_7. Click on the **Create** twisty and select **Document processing automations**.

### Business automations

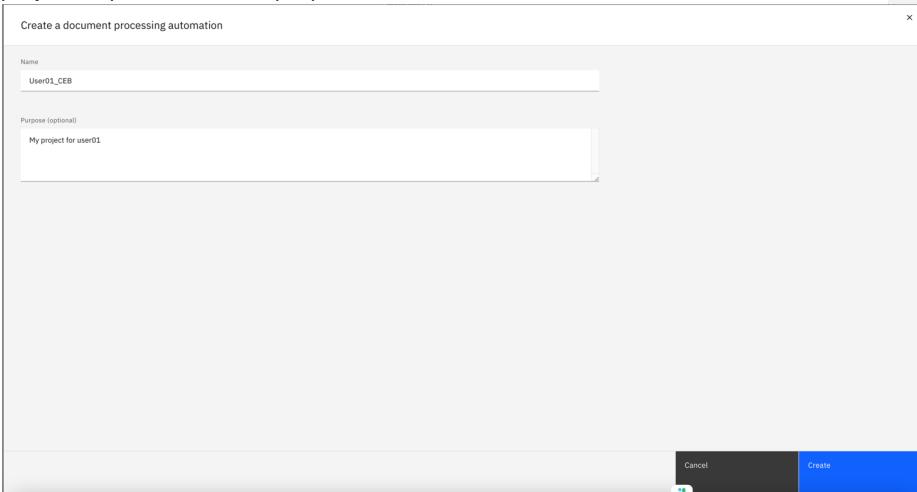
Create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way.

[Learn more](#)

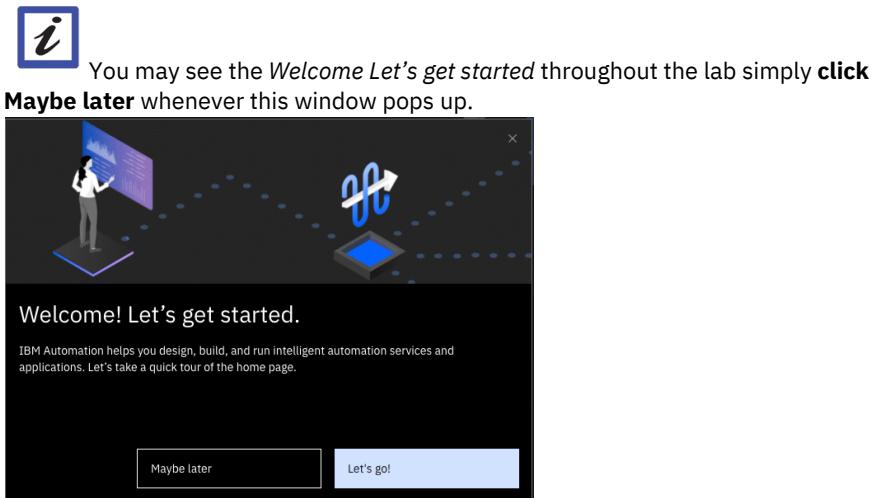
This screenshot shows the 'Create' dropdown menu expanded, revealing options for different types of automations. The 'Document processing automations' option is highlighted with a red box and a red arrow pointing to it from the previous screenshot. Other options listed in the dropdown include 'Decision automations', 'Workflow', and 'External'. To the right of the dropdown, there is a list of other automation categories: 'Document processing', 'Workflow', and 'External', each with an arrow icon.

## Automation Document Processing Lab

- \_8. In the Create a document processing automation window **enter a name** for the project. Optional, enter a purpose.



- \_9. Click on **Create** in the lower right-hand corner.



## 4.1 Reviewing the interface.

The screenshot shows the IBM Cloud Pak for Business Automation interface. At the top, there's a navigation bar with 'IBM Cloud Pak | Administration' and a search bar. Below it, the path 'Business automation / Clandis Baker Project' is shown. On the left, there are three tabs: 'Build' (selected), 'Enrich', and 'Configure'. The main content area is titled 'Document types and samples' and contains five sections: 'Classification model', 'Extraction model', 'Data standardization', and 'Document retention'. Each section has a status indicator (Ready or Not ready) and some numerical values. At the top right of the main content area, there are 'Share' and 'Version / Deploy' buttons. Below the main content, there's a status bar with 'Last shared | a few seconds ago' and 'Latest version | not yet Deployed | not yet'.

Upon opening the project, there are three major sections: **Build tab, Enrich tab, and Configure tab.**

On the top right, you find the *Share* and *Version/ Deploy* buttons.

This is a close-up view of the top right corner of the interface. It shows the 'Share' button and the 'Version / Deploy' button, both enclosed in blue-bordered boxes. Below each button is a status message: 'Last shared | 2 minutes ago' under 'Share' and 'Latest version | not yet' under 'Version / Deploy'. At the bottom, there's a 'Deployed | not yet' message.

The *Share* button is used to save your configuration to your GitHub repository.

The *Version / Deploy* button is used to create a snapshot, or version of your configuration. Like the *Share* button, the *Version* button will save your configuration, but will also create a version of it while retaining your previous version.

Once you have created a version of your configuration, you can also use this button to *Deploy* your version to the Business Applications area of ADP. You need to do this before you can go into the Business Application tile and configure your user interfaces.

### 4.1.1 Build Tab

This is what you will be spending most of your time on. The BUILD tab shows the guided configuration for building a Document Processing project. It shows the five steps required.

Document types and samples: Here you will define the document types that can be recognized by this automation and upload sample documents for training. By default, any project will be pre-populated with three pre-trained document types (Bill of Lading, Invoice, and Utility Bill).

Classification model: classification: Here we will teach the system how to recognize the different document types.

Extraction model: Here we will teach the system how to extract information for each document type based on the classification.

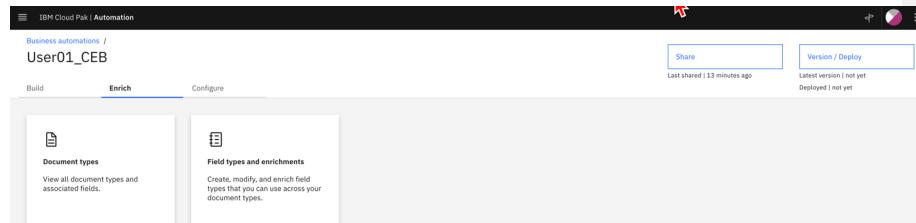
Data Standardization: This allows further refinement of the extracted information. For example, we want to standardize all dates to be formatted as YYYY/MM/DD. Having a standardized data format will help with any subsequent automation process.

Document retention: This allows us to define how long we want our documents to be kept in the system. Documents that have exceeded the retention period will be automatically expunged. This could be important for regulatory compliance or for managing the overall storage size.

#### 4.1.2 Enrich Tab

##### \_1. Click on the ENRICH tab.

Enrich provides a quick way to define your document types and the fields you wish to extract. In this section, we can define additional enrich rules. An example of an enrich rule is to specify the expected format for an invoice number (all numerical) or a driver's license. The more we can tell document processing about how different data will be formatted, the higher the chance it will recognize the information.



##### \_2. Click on FIELD TYPES AND ENRICHMENTS to begin. In this tile, you will see some of the pre-configured fields in the SYSTEM LIBRARY (sys). Customers can use

## Automation Document Processing Lab

these fields in their document type field definitions as needed.

Field type	Value type
Address block	String
Address information	Composite
Addressee	String
Boolean	Boolean
Building number	String
City	String
Country	String
Country code	String
Country name	String
Currency	Composite
CurrencyCode Object Type	String
Date	Date
Date Range	Composite
Decimal	Decimal

**Address block**

General	
Display name	Address block
Symbolic name	AddressBlock
Value type	String
Description	Address Block
Other possible names	
Required by default	Not required
Sensitive by default	Not sensitive
Validator	
Validators	4
Value format	
Text	
Extractors	None
Formatters	3
Converters	None

- \_3. Click on <your project name> in the bread crumb trail at the top to go back to the Build tab.

Business automations / Clandis Baker Project /

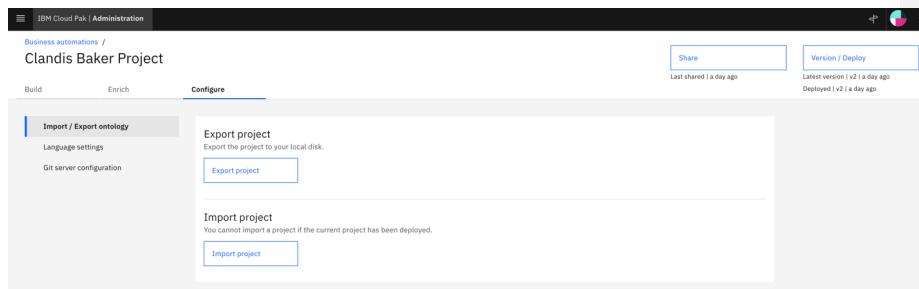
### Field types and enrichments

#### 4.1.3 Configure Tab

- \_4. Click on Configure Tab

This is where we can configure other operational aspects of the project. The export project creates a .zip file that contains the document types, field types and enrichments, which you can use to start training with new sample files. You can also decide to include the training model and the sample training files in your export if you want to move your entire project to a new instance of Document Processing for example. To import a project, select the .zip file to import. When you import a .zip file you have two options: overwrite the existing project or merge the existing project. If you merge the existing project, document types, field types, enrichments, and sample training files are imported unless there is a conflict. Models are not imported.

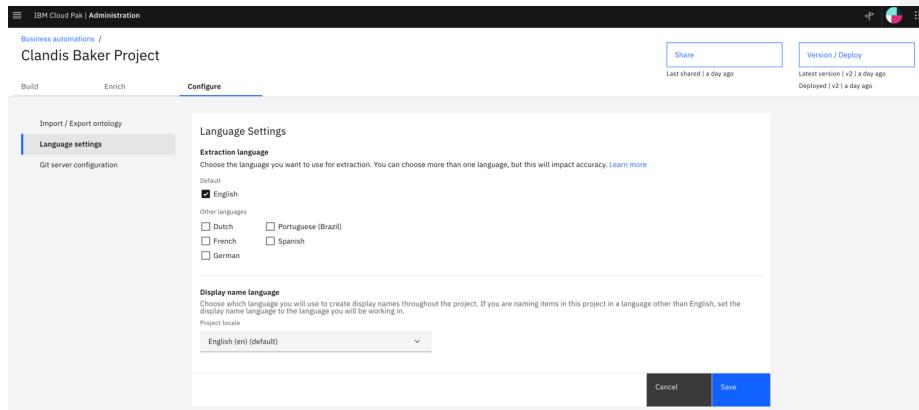
## Automation Document Processing Lab



In Extraction language, select which languages are used in the documents that you plan to process. You can choose English, Dutch, French, German, Brazilian Portuguese, or Spanish. Make sure to choose only the language or languages that are likely to be used in your document sets. Choosing more than one language can affect the accuracy of your document processing model.

In Display name language, select the language that you use to enter display names for fields and document types. These are the names that are displayed in the Designer and in the applications.

The display name language is also used in the Content Engine as the localized string locale setting for document classes and properties. Document Processing project deployment supports only one language per project. If your organization has multiple projects with different language settings, these projects cannot be deployed to the same Content Engine server if they share common properties. For example, when you define data definitions during data standardization, you cannot map a field to an existing data definition that was created in a different language.



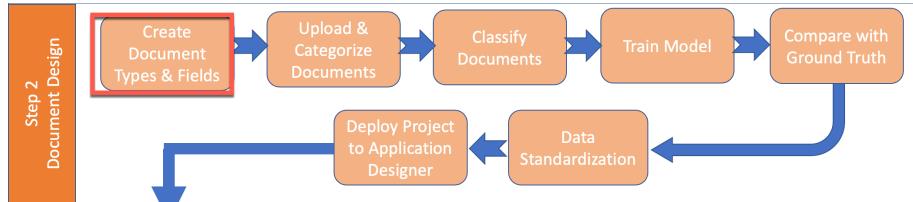
The Git server configuration is where you create a connection to the Git server for the first project that you create in Document Processing Designer. This setting applies to all

## Automation Document Processing Lab

subsequent projects that you create.

The screenshot shows the IBM Cloud Pak Administration interface. The top navigation bar includes 'IBM Cloud Pak | Administration', 'Business automation / Clandis Baker Project', and tabs for 'Build', 'Enrich', and 'Configure'. The 'Configure' tab is active. On the left, there are links for 'Import / Export ontology', 'Language settings', and 'Git server configuration'. The main content area is titled 'Git server configuration' and contains fields for 'Git vendor' (set to 'Gitea'), 'Git server organization URL' (set to 'https://icp4adeploy-gitea-svc:3000/content-designer'), 'Git server REST API URL' (set to 'https://icp4adeploy-gitea-svc:3000/api/v1'), 'Username' (set to 'git'), and 'Type of credentials' (radio button selected for 'API key'). Below these fields is a 'Credentials' section with a placeholder 'Enter a password or API key'. At the bottom are 'Test' and 'Save' buttons. To the right of the configuration panel, there are 'Share' and 'Version / Deploy' buttons, along with status indicators: 'Last shared 1 day ago', 'Latest version v2 1 day ago', and 'Deployed v2 1 day ago'.

## 5 Configure a Wage and Tax document type.



Before we use the guided configuration, you will configure some additional document types and fields used to extract data prior to uploading sample documents.

To do this lab, we will use the ENRICH tab to add fields to a newly created Wage and Tax document type.

### 5.1 Create Wage and Tax document type.

\_1. Click on **Build** tab to return to the start page.

Category	Status	Value	Details
Document types and samples	Ready	3 types	29 samples on average
Classification model	Ready	3 types trained	100% accuracy
Extraction model	Ready	3 types trained	97% accuracy
Data standardization	Not ready		
Document retention	Ready	3 types reviewed	

\_2. Click on the **ENRICH** tab

\_3. Click on **DOCUMENT TYPES**



We will now create a document type for Wage and Tax documents and fields to extract data from them.

\_4. Click on the **CREATE DOCUMENT TYPE** button in the top right corner.



- \_5. The Add document type window pops up. Enter **Wage and Tax** for the display name. There is no need to enter a symbolic name ADP will use the display name as a base. There's no need to add description in this lab unless you want to.

Add document type X

Display name 12/50  
**Wage and Tax**

This is the name that will show up for you in the system. You can use characters from any language.

Symbolic name 10/50  
**WageandTax**

This name will be used to identify the document type in the code.

Description (optional) 0/512  
Enter a description for this document type

**Fixed-format document type**  
Fixed-format documents have a fixed structure that remains the same for every document. Fixed-format documents types do not require as many sample documents to be trained in the extraction model.

This document type has a fixed format

Cancel Add

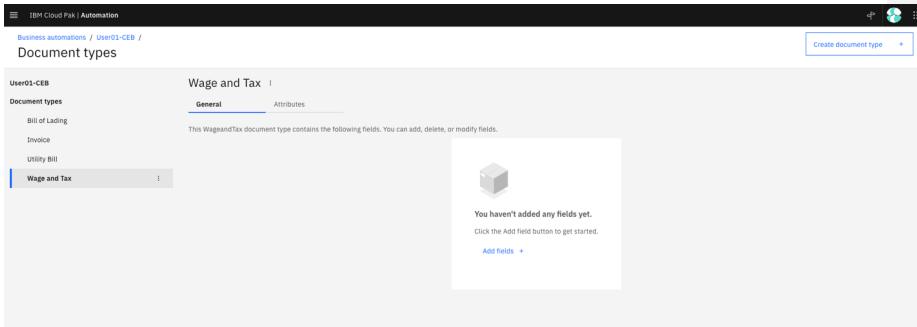


*Note: Notice the option for “Fixed-format document type”. If your form is static in nature or has a fixed structure that does not change, select this option so you will not have to provide as many samples. In our use case Wage and Tax documents have a variety of formats and are not static.*

- \_6. Click the **ADD** button.

You should now see your new document type (class) in the list of classes on the left.

## Automation Document Processing Lab

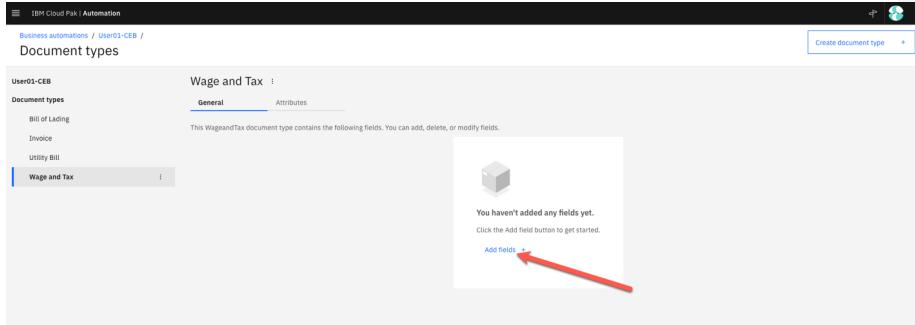


\_7. Select your **Wage and Tax doc type**. On the right, you should see an empty table of fields.

### 5.2 Create Field

We can now add some fields to the class. From examination of the forms, we can see there are different fields names, or they are not consistent across the forms. We'll need to add these different "aliases" during this process.

#### \_1. Click ADD FIELDS



\_2. Enter the following values under the GENERAL Settings header

## Automation Document Processing Lab

The screenshot shows the 'Create field' dialog in the IBM Cloud Pak | Automation interface. The 'General' tab is selected. The 'Display name' field contains 'Ex. Employee's name, Le nom de l'employé' and has a red border, indicating it is a required field. The 'Description (optional)' field is empty. The 'Symbolic name' field contains 'Enter a name'. The 'Field type' dropdown is set to 'sys:String'. The 'Aliases' field contains 'Enter an alternative name'.

- **Display Name:** **Federal Income Tax Withheld**
- **Field Type:**
  - **Sys:Decimal**
- **Is this field required:** **Yes**
- In Aliases enter other possible names. Case and punctuation are very import when creating aliases. Enter the alias listed below. These are representations of what it looks like on the different forms. **Press the “+” after entering each one or press Enter key:**
  - **2 Federal income tax withheld**
  - **2. Federal income tax**



*Note: the number two has a period after it*

You should now see the following:

The screenshot shows the 'Create field' dialog for the 'Wage and Tax' document type. The 'General' tab is selected. The 'Display name' field contains 'Federal Income Tax Withheld'. The 'Symbolic name' field contains 'FederalIncomeTaxWithheld'. The 'Field type' dropdown is set to 'sys:Decimal'. The 'Aliases' field contains '2 Federal income tax withheld' and '2. Federal income tax', separated by a plus sign. The 'This field is required' checkbox is checked.

- \_3. Click the **NEXT** button.



Field patterns are regular expressions that can be associated with a field to help identify and extract fields. A regular expression is a sequence of characters that define a search pattern. The use of regular expression patterns and extractors is optional. Regular expression patterns can provide extra information to potentially improve the accuracy in extracting the correct fields. Python syntax is used for defining the regular expressions. You will not be adding any field patterns in this lab.

- \_4. Click **NEXT** again on the Field patterns screen. You should now be on the **VALUE SETTINGS** page. This is where you can set up validators, formatters, and converters.

Commented [CB2]: Add convertor



Value Settings for a specific field; if the potential values follow a rule that can be expressed in a regular expression, you can specify an extractor. This pattern can match all the variations of your values. For example, the expected value for a Start Date field might be in a date format. You can create a regular expression pattern for `US Date` and then associate the extractor of `US Date` to your field.

Also, sometimes you want to extract a value that does not have a corresponding key in the document, but you know the pattern of the value. You can define the extractor and denote that the value might be anywhere in the document without attaching to the field name. This designation allows for the presence of a field name to be optional. For example, you want to extract the employee ID number, which can be described with a regular expression pattern. However, some documents show the employee number with a field name Employee ID, while other documents show the employee number without a corresponding field. You can specify the Extractor and be able to extract the employee ID number in both types of documents.

- \_5. The decimal data type can contain only integers to the left and right of a decimal point. But some of our data may contain commas between the integers and we only need two integers after the decimal point. Let's add a converter that will remove all extra punctuation and limit the number of integers after the decimal point to two. Click on the **Edit** button.

## Automation Document Processing Lab

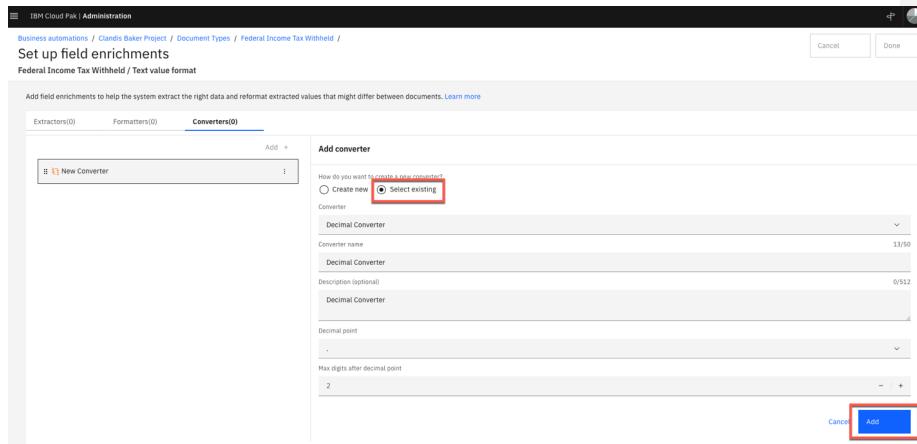
The screenshot shows the 'Value settings' section of the 'Value format' tab. It includes tabs for General, Field patterns, and Value settings. Under Value settings, there are sections for 'Value format' (Text, Extractors, Formatters, Converters), 'Value validators' (Low Confidence Validator, Datatype Mismatch Validator, Required Value Validator), and a 'Validator details' table.

### \_6. Click on Converters tab then click on blue button Add converter

The screenshot shows the 'Set up field enrichments' screen with the 'Converters' tab selected. It displays a table with columns for Extractors(0), Formatters(0), and Converters(0). A message at the bottom indicates 'No converters were added' and provides an 'Add converter' button.

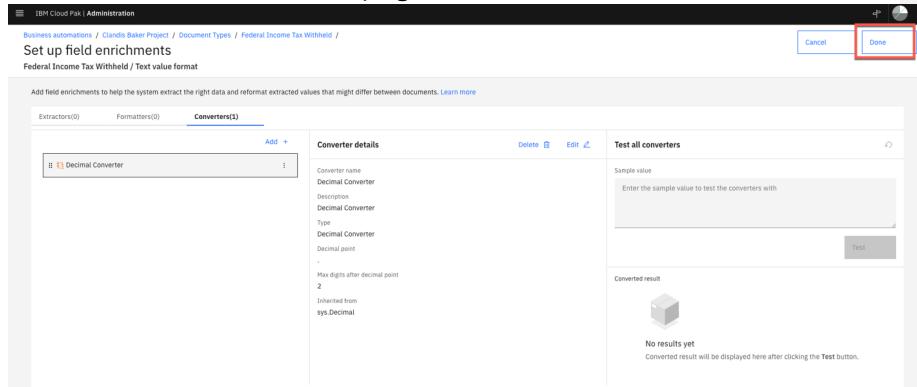
### \_7. You will be presented with the Add converter screen. Click on Select existing. This populates the converter name, description, Decimal point, and Max digits after decimal point for you. If you wanted to change the decimal point from a period to a comma you could do it here as they do in other countries outside the United States. Click the blue Add button.

## Automation Document Processing Lab



The screenshot shows the 'Set up field enrichments' interface in IBM Cloud Pak Administration. The 'Converters' tab is active. A 'New Converter' button is visible. A modal dialog titled 'Add converter' is open, showing the 'Select existing' radio button is selected. The 'Add' button in the dialog is highlighted with a red box.

- \_8. You will then be presented with Converter details information screen. On this screen you can also test your converters to make sure they are behaving like you intended. **Click on Done** at the top right.

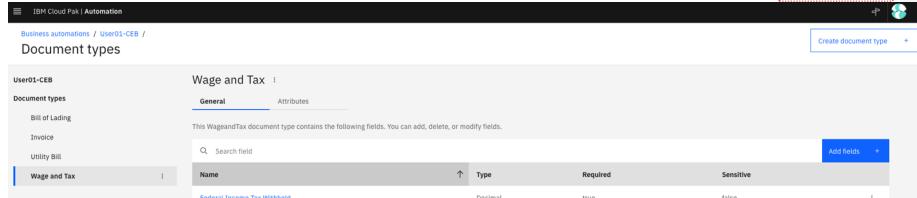


The screenshot shows the 'Converter details' screen for the 'Decimal Converter'. The 'Done' button is highlighted with a red box. The 'Test all converters' section shows a sample value input field and a 'Test' button, both currently empty.

## Automation Document Processing Lab

\_9. Click Create your screen should look like this with your first field created.

Commented [CB3]: This step may not be necessary after adding a converter.



Once it is created you will go back to the Document type page.

### 5.3 Create the Employee Name Address field.

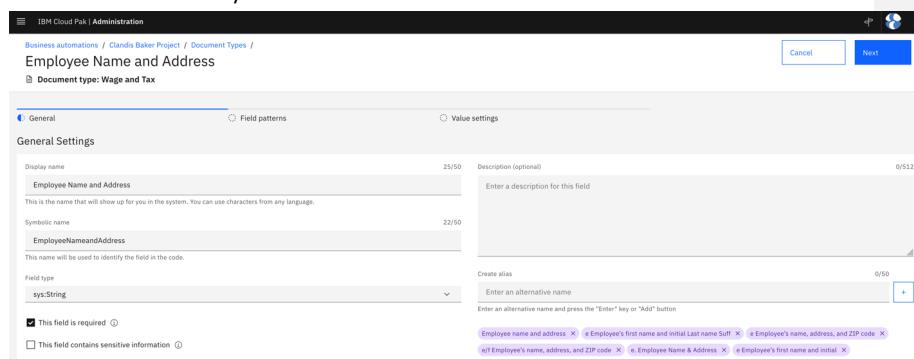
\_1. Click Add fields.

Give it the following parameters:

- Display name: **Employee Name and Address**
- Field Type = **sys:String**
- Required = **yes**
- Enter the following other possible names (aliases):
  - **Employee name and address**
  - **e Employee's first name and initial Last name Suff**
  - **e Employee's name, address, and ZIP code**
  - **e/f Employee's name, address, and ZIP code**
  - **e. Employee Name & Address**
  - **e Employee's first name and initial**

By default, the system will use the field name as an alias. So, you do not have to add it.

For example, below, Employee Name and Address (field name), would be automatically used as an alias even if you do not add it to the list



- \_2. **Click Next** no field patterns will be created.
- \_3. **Click Next** no value settings will be created.
- \_4. **Click Create** to finish creating the Employee Name and Address.

## 5.4 Create Employee Social Security Number Field

- \_1. **Click on ADD FIELDS**



Enter the following values in the GENERAL page.

- Display Name: **Employee Social Security Number**
- Field Type: **sys:Social Security Number**
- Is value required: **Yes**
- Other possible names (aliases). Remember, press RETURN or hit the '+' button on your keyboard between each entry:
  - **a Employee's social security number**
  - **a Employee's social security no.**
  - **a Employee's SSA number**
  - **a. Employee Social Security Number**
  - **Employee social security number**

Your screen should now look like the image below:

The screenshot shows the SAP Fiori interface for creating a new field. The 'General' tab is selected. The 'Display name' field contains 'Employee Social Security Number'. The 'Field type' dropdown is set to 'sys:Social Security Number'. The 'Is value required' checkbox is checked. In the 'Create alias' section, there are five entries: 'a Employee's social security number', 'a Employee's social security no.', 'a Employee's SSA number', 'a. Employee Social Security Number', and 'Employee social security number'. The 'Next' button is visible at the top right of the screen.

- \_2. **Click NEXT**
- \_3. **Click NEXT** again on the Field Patterns screen.
- \_4. **Click Create** on the Value settings.
- \_5. Create the following additional Fields.

The following table contains the values to use when adding the additional fields.

Follow the steps from the previous section to add the following fields. Don't forget to add your converter for datatypes of Sys:Decimal.

Display Name	Description	Type	Mandatory	Aliases
Employer Identification Number		sys:String	N	<ul style="list-style-type: none"><li>• b Employer identification number (EIN)</li><li>• b Employer's FED ID number</li><li>• b. Employer ID number</li><li>• Employer identification number</li></ul>
Employers Name and Address		sys:String	N	<ul style="list-style-type: none"><li>• c Employer's name, address, and ZIP code</li><li>• c Employer's Name &amp; Address</li><li>• Employers name and address</li></ul>
Social Security Wages		sys:Decimal	N	<ul style="list-style-type: none"><li>• Social security wages</li><li>• 3 Social security wages</li></ul>
Wages Tips Other Compensation		Sys:Decimal	N	<ul style="list-style-type: none"><li>• 1 Wages, tips, other compensation</li><li>• Wages, tips, other comp.</li><li>• 1 Wages, tips, other comp.</li><li>• 1. Wages tips, other comp</li><li>• Wages tips other compensation</li></ul>

Reference for various field types:



*Note: The basic default field types included in ADP are found here in the documentation*

## Automation Document Processing Lab

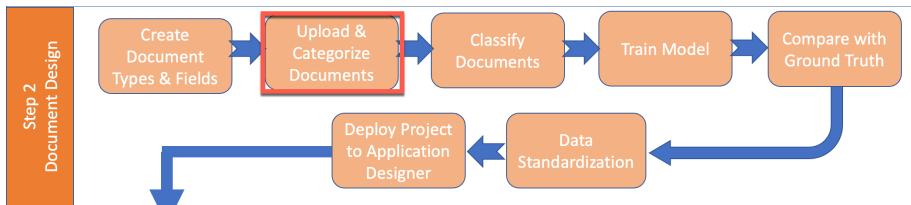
<https://www.ibm.com/docs/en/cloud-paks/cp-biz-automation/22.0.1?topic=enrichments-field-types-document-processing>

- \_6. Click on the <name of your project> in the breadcrumb link in the top left of your screen. In the following example the name of the project is <Clandis Baker Project>. This will take you back to the Enrich tab, Then Click the Build tab.

The screenshot shows the IBM Cloud Pak Administration interface. The top navigation bar includes 'IBM Cloud Pak | Administration' and a search bar. The breadcrumb path 'Business automation / Clandis Baker Project / Document types' is visible. On the right, there are buttons for 'Create document type' and a grid icon. The main content area is titled 'Wage and Tax' under 'Document types'. It has tabs for 'General' and 'Attributes'. A note states: 'This Wage and Tax document type contains the following fields. You can add, delete, or modify fields.' Below is a table of fields:

Name	Type	Required	Sensitive
Employee Name and Address	String	true	false
Employee Social Security Number	SocialSecurityNumber	true	false
Employer Identification Number	String	false	false
Employers Name and Address	String	false	false
Federal Income Tax Withheld	Decimal	false	false
Social Security Wages	Decimal	false	false
Wages Tips Other Compensation	Decimal	false	false

## 6 Document Types and Samples Overview



At this point in the process, we have created a new document type and configured the field names we want to extract off the document. For the system to know what to extract from your documents, it needs to be able to classify the documents. In this lab, we will teach the system to recognize the various document types on your system.

In the first part of the classification lab, you will explore the system's ability to automatically group similar documents together. This can be used to discover document types in a file share for example. You can also upload documents and have the system tell you what it finds. You would then use this information to create document types so you can classify the documents and data extract fields.

The project template comes pre-loaded with three document types: Bill of Lading, Invoice, and Utility Bill. In the last step we added a new document type Wages and Tax. In the BUILD tab of your project, you should now be seeing 4 document types. The three pre-loaded documents already have documents in them. You will be adding documents to the Wage and Tax document type. Your actual screen may vary from the following screen shot.

You will be asked to review the document categories the system finds and create the appropriate document types as needed.

### 6.1 Categorize documents.

For categorizing, we will have the system help us group similar documents together. To get started,

## Automation Document Processing Lab

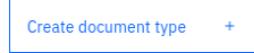
\_1. Click anywhere in the Document types and samples box.

The screenshot shows the IBM Cloud Pak Administration interface. At the top, it says "IBM Cloud Pak | Administration" and "Business automations / Clandis Baker Project". Below that, there are three tabs: "Build" (which is selected), "Enrich", and "Configure". On the right side, there are two buttons: "Share" and "Version / Deploy". Under "Build", there are four sections: "Document types and samples", "Classification model", "Extraction model", and "Data standardization". The "Document types and samples" section has a red box around it. It contains the text "Upload sample documents to define the types of documents you want the system to process." followed by a link icon. To its right, it shows "Ready" status, 4 types, and 22 samples on average. The other three sections show "Ready" status, 3 types trained, and 100% accuracy for the Classification model; 3 types trained and 97% accuracy for the Extraction model; and "Not ready" status for Data standardization.

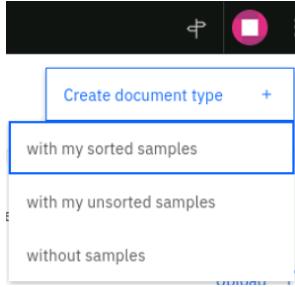
The CATEGORIZE feature analyzes each document and tries to find similarities between them. Based on these similarities, the system will divide the samples into categories for you to review. You can add documents or entire categories into either an existing document class or create new classes as needed.

Let's see what that looks like.

\_2. Click on **CREATE DOCUMENT TYPE** in the top right of the screen.



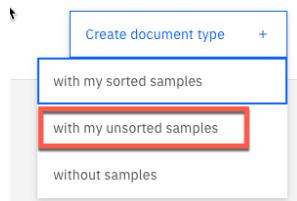
The drop down that appears:



If you have the same document types already separated into folders, you can choose the first option, *with my sorted samples*. The system would simply ingest the documents from each folder into a different group.

For this exercise, we will select the second option, *with my unsorted samples* and let the system sort the documents for us. Use this option when you don't know how many different document types there are.

**\_3. Select the second option titled **with my unsorted samples**.**



You should have already downloaded the files from [Section 2](#) to your laptop. You can select upload and grab all the files from where they were downloaded to on your laptop. Make sure you have already unzipped them.

**\_4. Click Upload to get document samples.**

From the downloaded sample documents open the folder name [Group 1 – Design Docs for Tax Lab.](#)

*Note: this will take several minutes, good time for some coffee or a stretch. Make sure to check ALL documents have been uploaded there are two pages or 12 items to verify.*

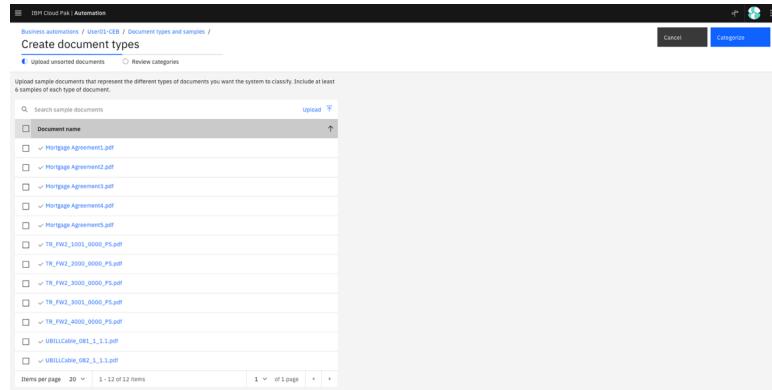
At the bottom of the window, you can select the number of items to display in the window or click on the arrows to move to the next page

A screenshot of a web-based document upload interface. At the top, there is a note: "Upload sample documents that represent the different types of documents you want the system to classify. Include at least 6 samples of each type of document." Below this is a search bar labeled "Search sample documents" and an "Upload" button with a file icon. A list of PDF files is displayed in a table format:

Document name
✓ Mortgage Agreement1.pdf
✓ Mortgage Agreement2.pdf
✓ Mortgage Agreements.pdf
✓ Mortgage Agreements.pdf
✓ Mortgage Agreements.pdf
✓ TR_FW2_1001_0009_P5.pdf
✓ TR_FW2_2000_0000_P5.pdf
✓ TR_FW2_3000_0000_P5.pdf
✓ TR_FW2_3001_0000_P5.pdf
✓ TR_FW2_4000_0000_P5.pdf
C URBLICache_0E1_1_1.pdf
C URBLICache_0E1_1_1.pdf

At the bottom of the interface, there are navigation controls: "Items per page: 20" with a dropdown arrow, "1 - 12 of 12 items", and a page navigation bar with arrows for "1" and "2".

\_5. Click on the blue CATEGORIZE button on the top right corner.



The screenshot shows the 'Create document types' interface. At the top, there's a navigation bar with 'IBM Cloud Pak | Automation' and a breadcrumb path: 'Business automation / Identity-GD / Document types and samples / Create document types'. Below the navigation is a section titled 'Upload sample documents that represent the different types of documents you want the system to classify. Include at least 6 samples of each type of document.' A search bar and an 'upload' button are present. A list of documents is shown, each with a checkbox and a checkmark icon. The list includes: Mortgage Agreement1.pdf, Mortgage Agreement2.pdf, Mortgage Agreement3.pdf, Mortgage Agreement4.pdf, Mortgage Agreement5.pdf, TH\_FW1\_1001\_0000\_PS.pdf, TH\_FW1\_2001\_0000\_PS.pdf, TH\_FW1\_3000\_0000\_PS.pdf, TH\_FW1\_3001\_0000\_PS.pdf, TH\_FW1\_4001\_0000\_PS.pdf, UBILICATE\_081\_1\_1.pdf, and UBILICATE\_082\_1\_1.pdf. At the bottom, there are dropdown menus for 'Items per page' (set to 20) and '1 - 12 of 12 items', along with a 'Cancel' and 'Categories' button.



*Note: The results may vary based on the documents uploaded, what the system already has learned, the version of ADP and more. Please look at this lab exercise from a high level. The categories you will be presented are the system's best guess on how they should be separated.*

You will need to:

- Review the categories to see if the documents were separated correctly.
- Move documents into either a NEW document type or into an EXISTING document type.
- There should be 3 types in the samples you were provided.
  - Wage and Tax
  - Utility bills
  - Mortgage Agreements
- You will need to assign either an entire category (i.e., all sample documents) or individual documents in each category to the Wage and Tax and Utility bills document types which already exist on your system.
- You will need to create a new document type for Mortgage Agreements.

After a few seconds, the system will mark the documents with a status of ready as seen in the above image.

## Automation Document Processing Lab

\_6. Click on each of the categories to see what was grouped together as shown below.

You can Click on any document to see a preview of it. This will help ensure the documents are correctly grouped.



*NOTE: The names of the files are not used in any way in this process. The files were merely named this way to make it easier for you to quickly ascertain whether the documents were grouped correctly.*

IBM Cloud Pak | Automation  
Business automation / User02-CER / Document types and samples / Create document types  
Upload unsent documents Review categories

Review each category, verify the documents, and assign each category to a new or pre-trained document type. Learn more

Categories (3)  
Category 1  
Category 2  
Category 3

Document types (4)  
Bill of Lading  
Invoice  
Utility Bill  
Wage and Tax

Category 1 sample documents (2)  
Document name  
UBILLCarve\_081\_1\_1\_1.pdf  
UBILLCarve\_082\_1\_1\_1.pdf

IBM Cloud Pak | Automation  
Business automation / User02-CER / Document types and samples / Create document types  
Upload unsent documents Review categories

Review each category, verify the documents, and assign each category to a new or pre-trained document type. Learn more

Categories (3)  
Category 1  
Category 2  
Category 3

Document types (4)  
Bill of Lading  
Invoice  
Utility Bill  
Wage and Tax

Category 2 sample documents (5)  
Document name  
Mortgage Agreement1.pdf  
Mortgage Agreement2.pdf  
Mortgage Agreement3.pdf  
Mortgage Agreement4.pdf  
Mortgage Agreement5.pdf

## Automation Document Processing Lab

The screenshot shows the 'Create document types' page. On the left, there's a sidebar with 'Categories (3)' and 'Document types (4)'. Under 'Categories', 'Category 3' is selected and highlighted with a red box. Under 'Document types', 'Bill of Lading' is selected and highlighted with a red box. The main area shows 'Category 3 sample documents (5)' with five PDF files listed. At the top right are 'Update categories', 'Back', and 'Finish' buttons.



*At the time of writing this documentation ADP was able to categorize the sample set into each category. This is not always the case, sometimes document types will be combined into one category, so it's very important to look at each category and verify documents.*

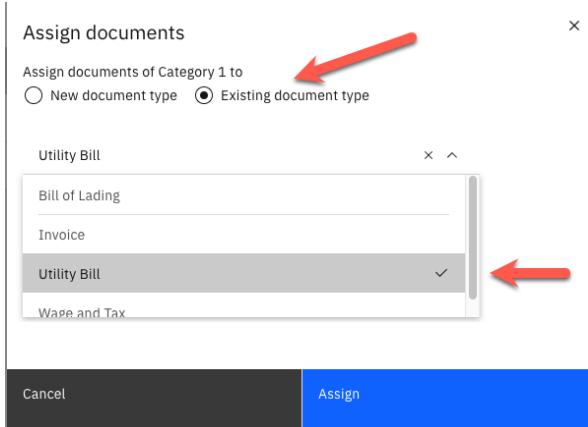
- \_7. If all documents within a category are correct as illustrated in the following screen shot, Click on the **3 dots** at the end of the category name.

The screenshot shows the 'Create document types' page. The 'Category 1' section is highlighted with a red box. The main area shows 'Category 1 sample documents (2)' with two PDF files listed. At the top right are 'Update categories', 'Back', and 'Finish' buttons.

- \_8. Select Assign to document type

The screenshot shows the 'Create document types' page. The 'Category 1' section has a red box around the 'Assign to document...' button. The main area shows 'Category 1 sample documents (2)' with two PDF files listed. At the top right are 'Update categories', 'Back', and 'Finish' buttons.

- \_9. Select Existing Document type then the appropriate document type from the drop-down list.

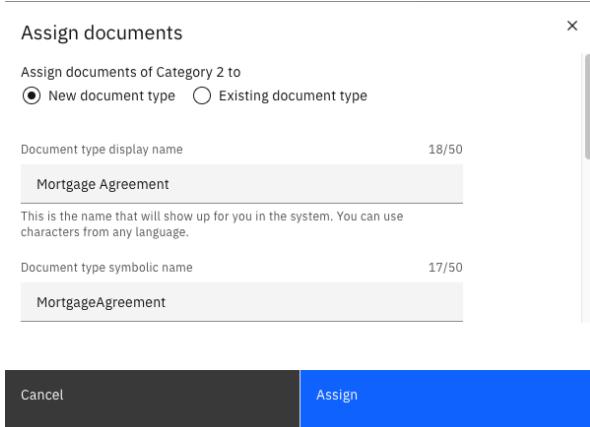


- \_10. Click Assign to close the dialog box.

- \_11. Select the next Category 2 and Click on the 3 dots and Select Assign to document type to a document class.

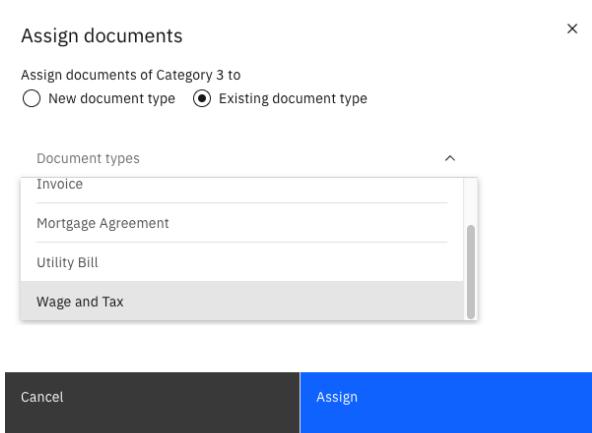
- \_12. This time Select a New Document Type. Since we have not defined a mortgage agreement document type yet.

- \_13. Enter Mortgage Agreement in the field



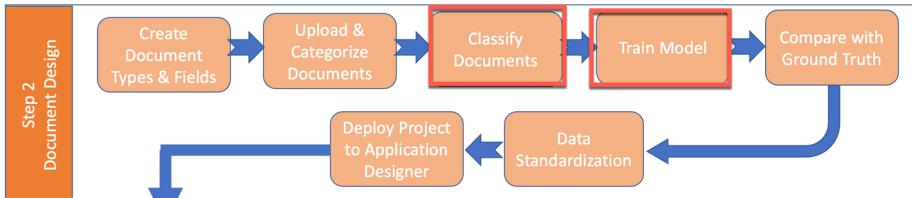
- \_14. Click Assign to have the system automatically rename and move the category into the Document Types section.

- \_15. Now for Category 3, **Click on 3 dots** and **Select Assign document type**.
- \_16. **Click Existing document type** and **Select Wage and Tax** from the drop down and then **Click on Assign**.



- \_17. Once you confirmed all documents are correctly classified into the correct document type, **Click Finish**

## 7 Train classification



Now that we have documents uploaded in the system, we are ready to train the classification. Note that although you don't need a ton of document samples to train (minimum of 5), you are going to get better accuracy if the system has a deeper understanding of the documents, so more could be better.

In this lab, we curated some documents samples for you. In normal circumstances, you would need to do this yourself. Make sure the documents you upload to train classification are good documents.

- Clean documents
- High resolution
- Representative of the document type(s)
- Accurately grouped and uploaded to Document Processing

This is NOT the time to try and trick the system. Uploading a document that doesn't recognize well would not help the system recognize the types of words, phrases, and concepts it needs to learn to classify documents correctly.

The most common error is introducing a sample document into the incorrect document type, usually by uploading them to the wrong document type. If that happens, you are introducing conflict into the classification. For example, an invoice added to Tax Forms may confuse the system and result in it thinking invoices are tax forms and vice versa. Once that happens, you need to clean your documents and retrain the system.

- \_1. Click on **<your project name>** in the cookie trail to return to the start page.
- \_2. Click anywhere in the **CLASSIFICATION MODEL** line

## Automation Document Processing Lab

The screenshot shows the IBM Cloud Pak Administration interface for the 'Clandis Baker Project'. The 'Build' tab is active. The 'Document types and samples' section shows 5 types and 20 samples on average. The 'Classification model' section is highlighted with a red box; it shows 3 types trained with 100% accuracy. The 'Extraction model' section shows 3 types trained with 97% accuracy. The 'Data standardization' section shows 'Not ready'. The 'Document retention' section shows 5 types reviewed.

Section	Status	Value
Document types and samples	Ready	5 types, 20 samples on average
Classification model	Ready	3 types trained, 100% accuracy
Extraction model	Retrain	3 types trained, 97% accuracy
Data standardization	Not ready	
Document retention	Ready	5 types reviewed

Once we open the classification model, we will be presented with details on how to perform the retraining. There are four basic steps – Confirm inputs, Review Samples, Review Training Results, and Test Trained model.

On the Confirm inputs screen here we can confirm all the documents that will be used in this training exercise. We can also use the opportunity to remove documents that are no longer relevant or upload additional documents.

- \_3. **Click Next** this will move from the Confirm inputs to the **Review Samples** step. Notice three documents have green icons next to them. These green icons show these documents have test samples already assigned. The new document types (Mortgage Agreement and Wage and Tax) do not have any test samples assigned yet therefore there's no green icons since we haven't assigned test sets yet.

## Automation Document Processing Lab

The screenshot shows the IBM Cloud Pak | Automation interface. At the top, it displays "Classification model" with an accuracy of 84.8%. Below this, there are tabs for "Confirm inputs", "Review samples", "Review training results", and "Test trained model (optional)". A message box at the top indicates changes were made since the last train, prompting to retrain for updated results. On the left, a sidebar lists "Document types": Bill of Lading, Utility Bill, Invoice, Wage and Tax, and Mortgage Agreement (selected). The main area shows "Mortgage Agreement sample documents (5)" in the "Training set" (100% total samples) and "Test set (0)" (0% total samples). The "Test set" message states "There are no documents in the test set. Include at least 1 document in the test set to view training results." There are arrows between the two columns to move documents.

- \_4. For the Mortgage Agreement move two documents to the Test set by **checking** and **click on the arrow** in between columns.

This screenshot shows the same interface after moving two documents from the training set to the test set. The "Test set" now contains 2 documents (40% total samples), while the "Training set" contains 3 documents (60% total samples). The message box still prompts to include at least one document in the test set to view training results.

- \_5. Select **Wage and Tax** on the Document types. This time let the ADP system Auto generate the 70/30 split to the test set. **Click Auto generate 70/30 split**

## Automation Document Processing Lab

The screenshot shows the 'Classification model' page in the IBM Cloud Pak for Automation interface. The top bar displays 'IBM Cloud Pak | Administration' and 'Classification model'. Below it, 'Accuracy' is shown as 100%. A message box indicates changes since the last train, with a note about document types and training/test sets. The 'Wage and Tax sample documents' section shows 5 samples in the 'Training set' and 0 in the 'Test set'. A red box highlights the 'Auto generate 70/30 split' button.



*The suggested split is 70/30 – that is, 70% of the available sample documents should be used for training, and we will validate the training results with 30% of the sample documents. This split is only a suggestion, and we can adjust it, but 70/30 is a good starting point.*

The screenshot shows the 'Classification model' page in the IBM Cloud Pak for Automation interface. The top bar displays 'IBM Cloud Pak | Administration' and 'Classification model'. Below it, 'Accuracy' is shown as 84.8%. A message box indicates changes since the last train. The 'Wage and Tax sample documents' section shows 5 samples in the 'Training set' and 2 in the 'Test set'. A red box highlights the 'Auto generate 70/30 split' button.

## Automation Document Processing Lab

- \_6. Click on **TRAIN** to launch the training. This may take a several minutes. You will see a progress bar has training progresses.



Once complete, you will be able to see the training results.



*What's happening: All the samples are run through multiple machine learning algorithms. These machine learning algorithms learn from the ground truth, the association between the sample documents (the OCR text) and the document types. The yielding models are then evaluated with the documents in test set. The model-predicted document types on these documents are compared with the human-provided answers to compute the accuracy. The top three accurate models are presented to the user, with the most accurate one being selected by default.*

You should see something like the following:

Document	Classified as	Classification result	Confidence
BOL_007.pdf	Bill of Lading	Correct	High
BOL_009.pdf	Bill of Lading	Correct	Medium
BOL_019.pdf	Bill of Lading	Correct	High
BOL_027.pdf	Bill of Lading	Correct	High
BOL_031.pdf	Bill of Lading	Correct	High
BOL_075.pdf	Bill of Lading	Correct	High

- \_7. Click on each of the document types. Notice the confidence levels. The both the Mortgage Agreement and Wage and Tax have a confidence of low. Low Confidence means we probably need to add more documents to our document class to get better confidence values.



You can easily see where the system may be struggling with Wage and Tax and Mortgage Agreement. You should look for document types that don't match the actual file or have a low confidence. Remember the more documents you give to train, the better the results.

\_8. **Click on Next.** This is the Test trained model. Here you can try and test other documents to see if they classified correctly. This step is optional but would be useful to try out the AI model to determine whether additional samples are necessary.

\_9. **Click Done**

## 7.1 How do I improve my results?

### 7.1.1 Option 1 – Add more samples.

To improve results, you would normally want to add more samples of the document ensuring they are clean and representative document to improve the system's understanding of the document.

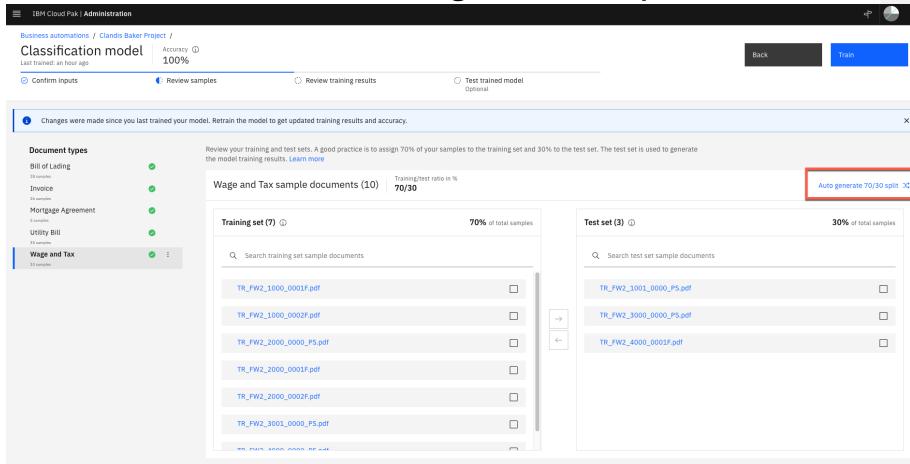
- \_1. **Click anywhere on Document Types and Samples.**
- \_2. **Click on Wage and Tax type.**
- \_3. **Click on Upload**

The screenshot shows the 'Document types and samples' section of the IBM Cloud Pak Administration interface. On the left, there's a sidebar with 'Document types' listed: Bill of Lading (1 sample), Invoice (1 sample), Mortgage Agreement (5 samples), Utility Bill (31 samples), and Wage and Tax (5 samples). The 'Wage and Tax' category is selected. The main area shows a list titled 'Wage and Tax sample documents (5)'. It includes a search bar, a table with columns for 'Document name' and 'Actions', and a list of PDF files: TR\_FW2\_1001\_0000\_PS.pdf, TR\_FW2\_2000\_0000\_PS.pdf, TR\_FW2\_3000\_0000\_PS.pdf, TR\_FW2\_3001\_0000\_PS.pdf, and TR\_FW2\_4000\_0000\_PS.pdf. An 'Upload' button with a file icon is highlighted with a red box at the top right of the list area.

- \_4. From the zip files you downloaded and unzipped earlier upload all the files from the directory *Group 2 - Classification Results Increase Set*.
- \_5. Go back to the **Build** tab then let's retrain the Classification Module again.
- \_6. **Click anywhere on Classification model.**
- \_7. **Click on Wage and Tax.**

## Automation Document Processing Lab

\_8. Click Next button. Also click on the Auto generate 70/30 split.



\_9. Click Train button.

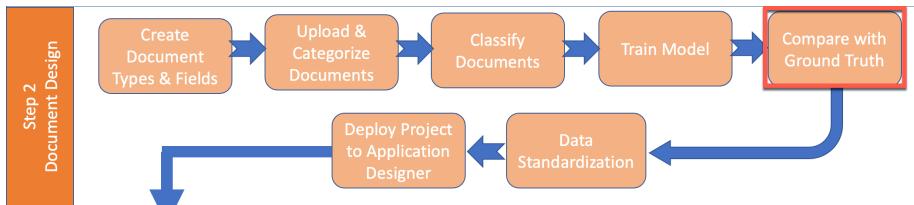
\_10. Now look at the confidence score for Wage and Tax.

\_11. Click Next and then Click Done

### 7.1.2 Option 2 – Review all uploaded samples.

- remove those that are not a clear representation.
- remove those that are poor quality documents.
- carefully confirm that none of the samples contain multiple document types in the file. This is a common occurrence. A document is listed as a Purchase Order, but in the back pages, also contains other document types in that same file. This confuses the system.

## 8 Data extraction



At this point, we have defined a document type, told the system which fields we want off the document and trained the system on how to recognize (classify) the document. In the Data Extraction portion of the lab, we will upload new Wage and Tax documents to Document Processing and see how our earlier configuration of the document type and related fields are working. This is comparing a new document extracted elements with the ground truth. Once we open Extraction model, we will be presented with details on how to perform the retraining. There are five basic steps – Review samples, Add fields, Teach the model, Review the trained model, and Test the model.

- From the guided configuration screen, **Click** anywhere in the **Extraction model** box.



*Note: the status will reset to Retrain if it detects something may have changed. This is just a reminder that if you indeed changed something, you may benefit from retraining the model.*



- Next **Click** on the **Wage and Tax** document type under the Document Types section.

Like in the classification step, ADP needs to have the documents divided into a training and test sets. In general, *deep learning*-based AI requires a larger number of sample documents to achieve a reasonable result. But since our environment does not have GPU, deep learning is not turned on.

## Automation Document Processing Lab

You should have something that looks like what you see in the following screen shot.

The screenshot shows the 'Extraction model' step in the 'Business automation' project 'Claudia Baker Project'. The 'Wage and Tax' document type is selected. A yellow warning box at the top right says: 'Please make sure you have at least 1 reviewed document to train the model. Review your training and test sets. A good practice is to assign 70% of your samples to the training set and 30% to the test set. The test set is used to generate the model training results. Learn more'.

The interface shows two lists of documents:

- Training set (6)**: 60% of total samples. Includes files: TR\_FW2\_1000\_0002F.pdf, TR\_FW2\_1001\_0000\_P5.pdf, TR\_FW2\_2000\_0002F.pdf, TR\_FW2\_3001\_0000\_P3.pdf, TR\_FW2\_4000\_0000\_P5.pdf, TR\_FW2\_4000\_0001F.pdf.
- Test set (4)**: 40% of total samples. Includes files: TR\_FW2\_1000\_0001F.pdf, TR\_FW2\_2000\_0000\_P5.pdf, TR\_FW2\_3000\_0002F.pdf, TR\_FW2\_3000\_0000\_P5.pdf.

At the bottom right, there is a button labeled 'Auto generate 70/30 split'.

\_3. Again, lets train with a 70/30 spilt. **Click Auto generate 70/30 split.**

\_4. **Click on the NEXT button at the top.**



You will now be on the Add fields bread crumb. If there were more fields to add we could do it here. But since we have already added all the fields needed, proceed to the next step.

\_5. **Click the Next button.** You are now at the "Teach model" bread crumb.

Teach the model is where you will spend most of your time. We can see that our documents are "not ready", so we'll need to teach the model with new documents.

## Automation Document Processing Lab

### \_6. Click on Teach Samples.

The screenshot shows the 'Extraction model' screen in the IBM Cloud Pak for Integration interface. The top navigation bar includes 'IBM Cloud Pak | Administration', 'Business automations / Claudia Baker Project /', 'Back', 'Train model', and a three-dot menu. Below the navigation is a toolbar with buttons for 'Review samples', 'Add fields', 'Teach model' (which is highlighted in blue), 'Review training results', and 'Test model'. A dropdown menu for 'Optional' is also present. On the left, a sidebar lists 'Document types' such as Bill of Lading, Invoice, Mortgage Agreement, Utility Bill, and Wage and Tax, with 'Wage and Tax' selected. The main area displays a message: 'Please make sure you have at least 1 reviewed document to train the model.' followed by 'Wage and Tax sample documents (7)'. Below this is a table with columns: Document name, Status, Fields reviewed, and Date added. The table lists seven PDF files, all marked as 'Not ready' with 0/7 fields reviewed, all added on April 3, 2023, at 12:27 PM. At the bottom right of the main area is a red box highlighting the 'Teach samples' button.



*Note: Your individual results may vary based on the exact documents you upload, how you configure your fields etc. Therefore, general guidance is given here versus exact step by step instructions.*

### \_7. We will now review the fields that were extracted, correct any that may be wrong and add others.

You should now see the field data extracted by the system. Nothing has been trained yet. All it is doing is using the field name and aliases we entered when we created the document class to locate data. Now, you need to correct and improve the model.

## Automation Document Processing Lab

IBM Cloud Pak | Administration

TR\_FW2\_1001\_0000\_P5.pdf | Not ready

Show detected fields Keyboard shortcuts on

Match data underlined in blue to the selected field or draw your own boxes around data in the document.

Sort by: Date created

Field Name	Value Captured
Federal Income Tax Withheld	<input type="text"/> Text 123456789.99
Field label (optional)	<input type="text"/> Captured field label 123456789.99
Field value	<input type="text"/> Captured field value 123456789.99

Pending alias | View all aliases (3)  
None ()

Save selection

Employee Name and Address  
Required

Employee Social Security Number  
Required

Employer Identification Number  
Text

Employers Name and Address  
Text

Social Security Wages  
Text

Mark this document as ready for training. (1)

Previous sample Next sample

Form W-2 Wage and Tax Statement 2020 Department of the Treasury—Internal Revenue Service  
Copy 1—for State, City, or Local Tax Department

22222 Employee's social security number OMB No. 1545-0008  
577-22-3048  
14-023285  
e Employer's name, address, and ZIP code  
Test and Run Inc.  
5432 Aldrich Rd  
Minneapolis, MN 55411  
d Control number 210220-A13  
f Employee's address and ZIP code  
15 State Employee's state ID number MN  
16 State wages, tips, etc. 795037  
17 State income tax 18000.00  
18 Local wages, tips, etc. 1260.00  
19 Local income tax 17700.00  
20 Local rate 500.00  
21 Mpls



Note: You may see different results than shown on the image above. Depending on how the algorithms interpreted the results you could see either type of extraction.

IBM Cloud Pak | Administration

TR\_FW2\_1000\_0001F.pdf | Not ready

Show detected fields Keyboard shortcuts on

Match data underlined in blue to the selected field or draw your own boxes around data in the document.

Sort by: Date created

Field Name	Value Captured
Federal Income Tax Withheld	<input type="text"/> Text 1800.00
Field label (optional)	<input type="text"/> Captured field label 1800.00
Field value	<input type="text"/> Captured field value 1800.00

Recommended matches (1)  
Matches are ranked in order of confidence. Choose one and save or dismiss to draw your own.

Field label Field value  
Federal income tax 1800.00

2 Federal income tax 1800.00 withheld

Edit selection Dismiss Seeing duplicates? (1)

Pending alias | View all aliases (3)  
Detected alias already exists (1)

Save selection

Employee Name and A... Required

Employee Social Secur... Required

Form W-2 Wage and Tax Statement 2020 Department of the Treasury—Internal Revenue Service  
Copy 1—for State, City, or Local Tax Department

22222 Employee's social security number OMB No. 1545-0008  
577-22-3048  
14-023285  
e Employer's name, address, and ZIP code  
Test and Run Inc.  
5432 Aldrich Rd  
Minneapolis, MN 55411  
d Control number 210220-A13  
f Employee's address and ZIP code  
15 State Employee's state ID number MN  
16 State wages, tips, etc. 795037  
17 State income tax 18000.00  
18 Local wages, tips, etc. 1260.00  
19 Local income tax 17700.00  
20 Local rate 500.00  
21 Mpls

Let's spend some time showing how to go about correcting these issues to help the system learn how to extract the values accurately.

## 8.1 Correcting extracted values

Let's start with the Federal Income Tax withheld field (i.e., The first one in the 'Fields to extract' list). Again, you may see different results based on your forms and how the different algorithms behaved on that particular document during extraction.

- \_1. ADP may have already preselected the first field like in the first screen shot below. But ADP can also show the characters it recognized on the page with blue lines (second screen shot below) If your result is like the first screen shot then **Click** blue button **Save section**. Otherwise, if you got blue lines **Click** on the **number** below the heading "**Federal Income tax withheld**" in the image.

The image consists of two side-by-side screenshots from the IBM Cloud Pak for ADP interface. The left screenshot shows a W-2 Wage and Tax Statement form for the year 2020. The right screenshot shows a 'Value Captured' panel for the 'Federal Income Tax Withheld' field. The panel lists several recommended matches, with the top one being 'Federal income tax withheld 1800.00'. A red box highlights the 'Save selection' button at the bottom of the panel.

## Automation Document Processing Lab

The screenshot shows the IBM Cloud Pak Administration interface with the document **TR\_FW2\_1001\_0000\_PS.pdf**. The document is a **W-2 Wage and Tax Statement** for the year **2020**. A pop-up window titled "Save Selection" is open, listing captured field values:

Field Name	Value Captured
Federal Income Tax Withheld	123456789.99
Social security tax withheld	123456789.99
Medicare tax withheld	123456789.99
Allocated tips	123456789.99
Dependent care benefits	123456789.99

The "Field label (optional)" section contains "Federal income tax withheld" with a "Draw" button and a "Captured field label" input field.

- \_2. Again, depending on your specific results. If ADP was able to find the field and will ask if you want to save match of value captured along with the field label. **Select Save Selection**. Otherwise, If your results were the recognized characters with blue lines then in the pop up window that comes up **select Save match**

The screenshot shows the IBM Cloud Pak Administration interface with the document **TR\_FW2\_1001\_0000\_PS.pdf**. The document is a **W-2 Wage and Tax Statement** for the year **2020**. A pop-up window titled "Save Selection" is open, listing captured field values. A red box highlights the "Save match" button in the "Field label (optional)" section.

Field Name	Value Captured
Federal Income Tax Withheld	2 Federal income tax withheld
Social security tax withheld	123456789.99
Medicare tax withheld	123456789.99
Allocated tips	123456789.99
Dependent care benefits	123456789.99

Notice a green check mark signifies this field is complete.

## Automation Document Processing Lab

The screenshot shows the IBM Cloud Pak Administration interface. On the left, a W-2 Wage and Tax Statement form is displayed. On the right, a data capture panel titled "Match data underlined in blue to the selected field or draw your own boxes around data in the document." is shown. The panel lists various fields from the W-2 form and their corresponding captured values. A green checkmark is present next to the value for "Federal Income Tax Withheld". Three ellipses (...) are located next to this checkmark, indicating options to clear or update the data.

Field Name	Value Captured
a Employee's social security number	123456789.99
b Employer identification number (EIN)	123456789.99
c Employee's name and ZIP code	Long Lengthy The Corporation 56334 Full Sized Avenue Unit 1234 Minneapolis, Minnesota 55411-1234
d Control number	123456789.99
e Employee's first name and initial	Michael Robert David Smithson III
f Employee's address and ZIP code	56334 Full Sized Avenue Unit 1234 Minneapolis, Minnesota 55411-1234
18 State Employee's state ID number	MN
18 State wages, tips, etc.	123456789.99
17 State Income tax	123456789.99
18 Local wages, tips, etc.	123456789.99
19 Local Income tax	123456789.99
20 Locality rate	MPLS
<b>W-2 Wage and Tax Statement</b>	<b>2020</b>
Department of the Treasury—Internal Revenue Service	

The 3 ellipses next the green check mark allow you to clear the data or update ADP to there is no field with this data in the current view.

- \_3. Move to Employee Name and Address field by clicking in the grey area on that field name. In our two possible outcomes depending on the algorithms. ADP did pick up the address but missed the name. Or the algorithm may have picked up the address and not the name. Or it may have gotten the correct field.

If the field is not correct **Click** on the **Dismiss** button.

Now under the Field label **select Draw** button and using your mouse grab or lasso around “Employee’s first name and initial”.

This screenshot shows the same W-2 form and data capture panel as the previous one. However, the "e Employee's first name and initial" field has been highlighted with a green box. A "Recommended matches" panel on the right side of the data capture panel shows a single result: "4326 Aldrich Rd Minneapolis, MN 55412". Below this, the "Field label" section shows the same field name again, with a "Draw" button next to it. The "Field value" section shows the captured value "4326 Aldrich Rd Minneapolis, MN 55412".

If you got the blue lines, you would notice that only the “e Employee’s first name and initial” have blue marks. In this case the values for name and address where not located. Using Draw button and using your mouse grab or lasso around “Employee’s first name and initial”.

- \_4. We are interested in getting the “Employee’s First Name” data and address for the field value. **Click** on the **Draw** button under Field value. Using your mouse select the appropriate values for Name and address (green box), then **Click Save selection**

The screenshot shows the IBM Cloud Pak Administration interface with the "Field Value Capture" tool open over a PDF of a W-2 form. The PDF displays various tax-related fields like wages, taxes withheld, and employee information. The "Field Value Capture" tool on the right lists captured fields such as "Federal Income Tax Withheld" and "Employee Name and Address". Two specific fields, "e Employee's first name and initial" and "d Employee's address and ZIP code", are highlighted with red boxes and have the "Draw" button applied. A green box highlights the "Save selection" button at the bottom of the capture tool. The W-2 form at the bottom shows fields like Social Security number (22222), Employee identification number (577-22-3048), and various tax amounts.

- \_5. For the Employee Social Security field if it looks good, **Click on Save selection**. Or if the blue lines are present instead **select** the value displayed to populate the field and **Click Save match** then **Click on Save selection**.
- \_6. Continue to process for the remaining fields, using either method as described above, clicking on the Save selection if ADP picked up the correct field label and field value or select the blue line values to populate both the field label and field value or finally if both fields are wrong use the *Dismiss* and use blue lines if Key Value Pair (KVP) is correct or drawing a box around needed label or value.

## Automation Document Processing Lab

- \_7. Once complete **check the box** next to “Mark this document as ready for training” at the bottom

The screenshot shows the IBM Cloud Pak Administration interface for document processing. On the left, a W-2 Wage and Tax Statement for the year 2020 is displayed. The form includes fields for employee information (Social Security Number: 22222, Employer Identification Number: 14-023285), wages (Total Wages: 18000.00), taxes (Federal Income Tax Withheld: 1800.00), and other details like Medicare and tips. On the right, the extracted data is shown in a structured format with dropdown menus for field labeling. A red arrow points to the checkbox labeled "Mark this document as ready for training" located at the bottom of the extraction interface.

- !** \_8. Review **ALL** other fields carefully. **Do not leave any incorrect values.** You can adjust or delete values as needed by clicking on Edit selection. If you leave incorrect values, the system will assume they are correct and LEARN them as if they were good values.

- \_9. Repeat steps for Next Sample

Over the course of next few samples you may find that ADP has extracted the wrong results, perhaps getting a value that is above when it should have been below. If this is the case and you pick you a blue underline data, but the results are wrong. Simply use the draw box for the Field Label and Field Value.



*Note: When completing the remaining documents, you may run across ADP finding the fields but perhaps on the second image or third image on the page. Try to keep all Key Value Pairs (KVP) on the same image.*

## Automation Document Processing Lab

\_10. Once complete review of all the sample documents **Click on the Back link**

The screenshot shows the JD Edwards Cloud Pack Administration interface. At the top, there are three tabs: 'TR\_FW2\_4000\_0005F.pdf' (highlighted with a red box), 'Form W2 Wage and Tax Statement', and 'Form W2 Wage and Tax Statement'. Below the tabs, there is a search bar and a 'Show detected fields' checkbox. The main content area displays three sample documents. The first document, 'TR\_FW2\_4000\_0005F.pdf', has a red box around its 'Back' button. The second and third documents are 'Form W2 Wage and Tax Statement' forms. The third form has a blue box around its 'Wages tips, other comp.' field.

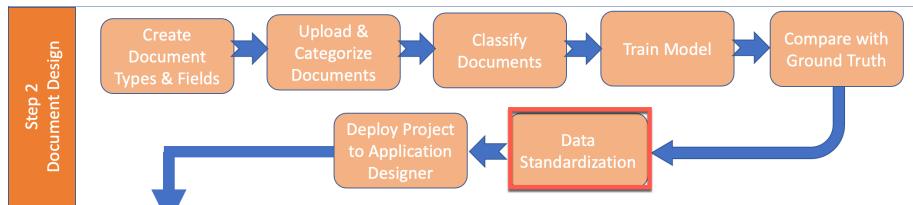
## 8.2 Train extraction model

We will be performing the quick training in this lab due not having a GPU in our TechZone architecture. A GPU is only needed a development environment and is not needed in either a production or runtime environment. The Deep Learning capabilities have been disabled on this training environment. You can find instructions in the Appendix for when you have access to a server with it enabled.

\_1. Click Train button.

This will take several minutes. (Good time for a break)

## 9 Data standardization



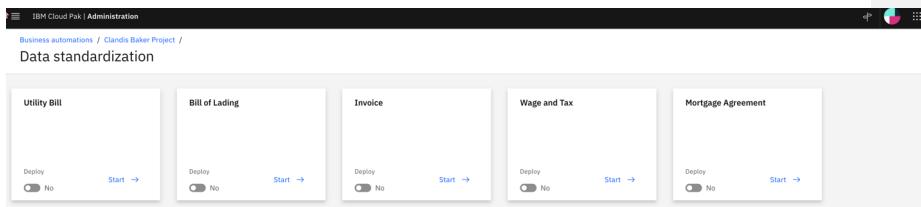
Next, we may need to standardize the data that will be presented in the user interface and how it will be stored in the FileNet repository for example. Data standardization is the process of defining attributes for a data field in a standardized way. This is done using data definitions. These definitions can be used across projects, and across different applications within the Cloud Pak for Automation. Each data definition has a title, description, and a datatype. We can also set a data definition as required or not. When a document is ingested into ADP, it results in a list of Key Value Pairs' (KVP) for that document. The Designer maps some of these KVP's to fields and teaches the model on how to extract the fields from the full list of KVP's. The designer then maps some of those fields to data definitions for a particular document type. Only the fields that have been mapped to data definitions will become Content Process Engine properties.

1. Return to the guided configuration flow and **Click** anywhere in the **Data standardization** box

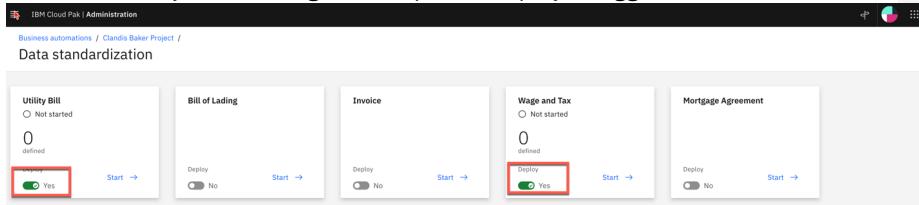
Document type status	Count	Accuracy
Ready	4 types 22 samples on average	
Ready	3 types trained 100% accuracy	
Ready	3 types trained 97% accuracy	
Not ready	Data standardization Map fields to new or existing data definitions.	

Here, you will see a list of available document types. Only the ones which have Deployed turned on will be visible in the verify interface and will have fields stored in FileNet.

## Automation Document Processing Lab



\_2. Ensure the Utility Bill and Wages and Tips and Deploy is toggled to Yes

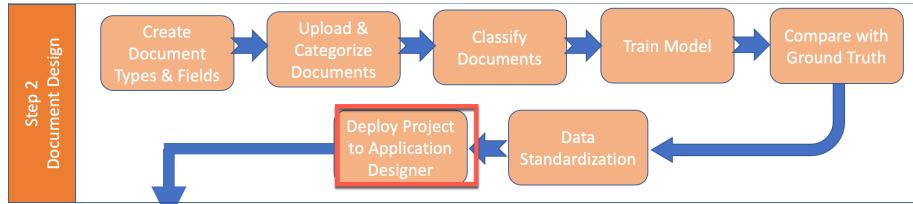


\_3. Click on Start on either selected deployment.

This is where we begin defining the data field attribute definitions. You could create a new data definition and configure them. We will NOT be creating/defining any data fields for this lab.

\_4. Return to the guided configuration screen by Clicking on <your project> name at the top of the screen.

## 10 Version and deploy your project

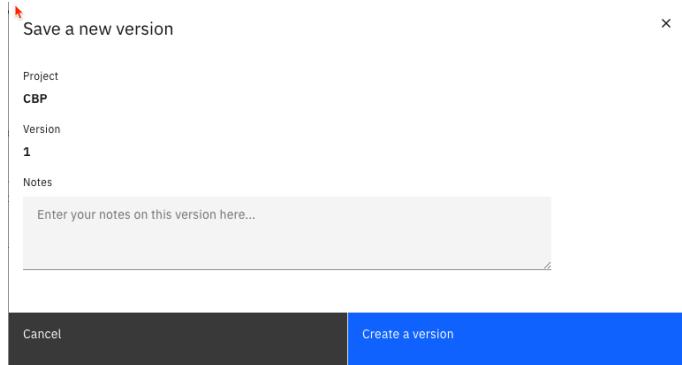


At this point in our Designer project, we have defined a document type, labeled the fields we want from the document, trained (classified) the system to recognize the document type, reviewed the extracted fields we wanted and standardized (mapped) the document fields to our output.

Now that we completed the configuration of the content extraction project, we need to save and deploy the design project to the application side. This will allow you to test your project using a client runtime interface.

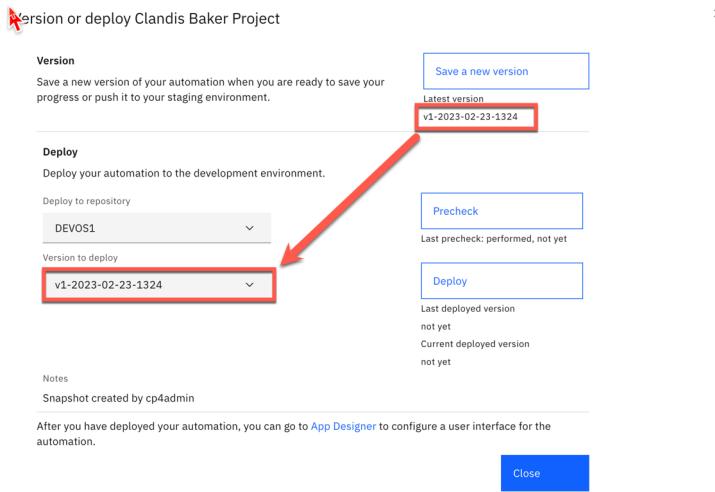
\_1. If not already there, return to the guided home screen by clicking on your project name. Then **Click Version / Deploy**

\_2. Click **Save a new version**. A Save a new version window pops up.



\_3. **Click on Create a version.**

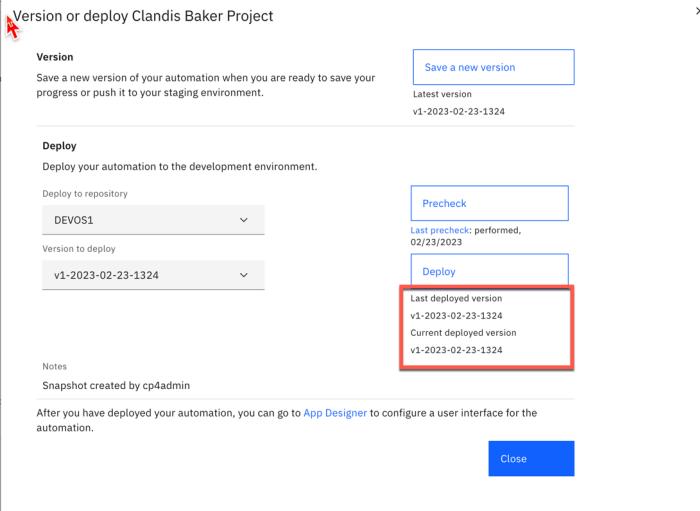
- \_4. Once the version is saved, you should see the version in the Version to deploy drop down list



... also, in the top corner has the “Latest Version.”

- \_5. Click on the Deploy button. This will also take a minute or two to deploy.

Once completed, you should have a notice that the project was deployed.





Note that you do not have to remain in the deploy screen while it is versioning or deploying. You can always click the button and then go back into any other screen if you like. It will run in the background. If you do this, just keep an eye on the top right of your screen for deployment status.

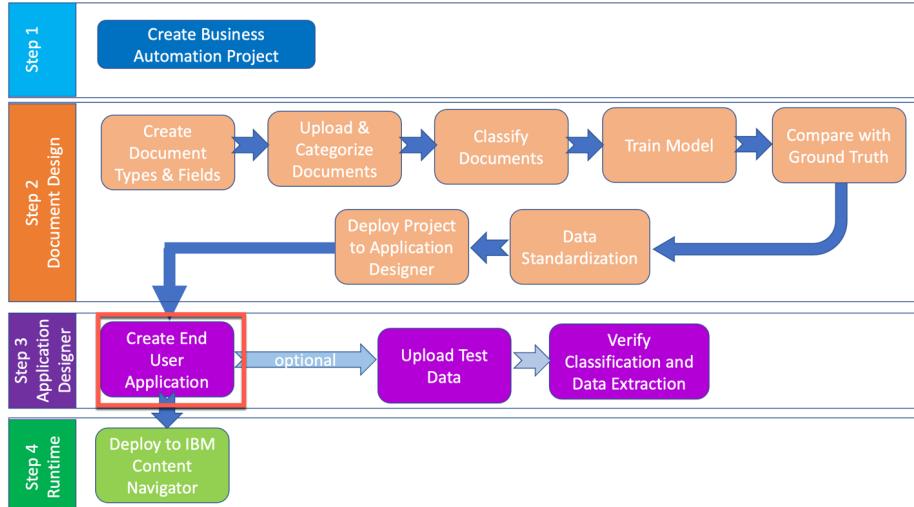
**\_6. Click Close button.**

Once deployed, proceed to the next steps.

From the home screen you can see the latest version and deployment

The screenshot shows the ISH Cloud Pak administration interface for the 'Clandis Baker Project'. The main area displays several configuration sections: 'Document types and samples' (Ready, 5 types, 20 samples on average), 'Classification model' (Ready, 5 types trained, 100% accuracy), 'Extraction model' (Ready, 4 types trained), 'Data standardization' (Not ready, 0 types reviewed), and 'Document retention' (Ready, 5 types reviewed). In the top right corner, there are two buttons: 'Share' and 'Version / Deploy'. The 'Version / Deploy' button is highlighted with a red box. Above these buttons, the text 'Last checked 1 minute ago' is displayed. Below the 'Version / Deploy' button, it says 'Latest version | v1 | 13 minutes ago' and 'Deployed | v1 | 4 minutes ago'.

## 11 Application designer



At this point we have designed or built a project that consists of document types, data or file types and methods to extract the desired data. The next major section of this lab is to build the user interface using the Application Designer. IBM provides two application templates for Document Processing

1. Batch Document Processing template – used to process batches of documents.
2. Document Processing Template – used to process single documents.

The lab will have you create a new batch processing application. We will quickly explore the various tabs in the interface, preview what the IBM Content Navigator (ICN) client would look like using the Preview feature and then publish our application to ICN where we will process a batch of documents.

Changes to the application itself will not be in the scope of this lab.

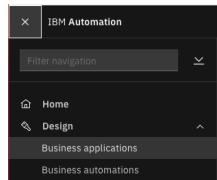
### 11.1 Create your Runtime Application.

- \_1. Return to the starting screen by **clicking the hamburger** in the top left.



and selecting **Business Applications**

## Automation Document Processing Lab



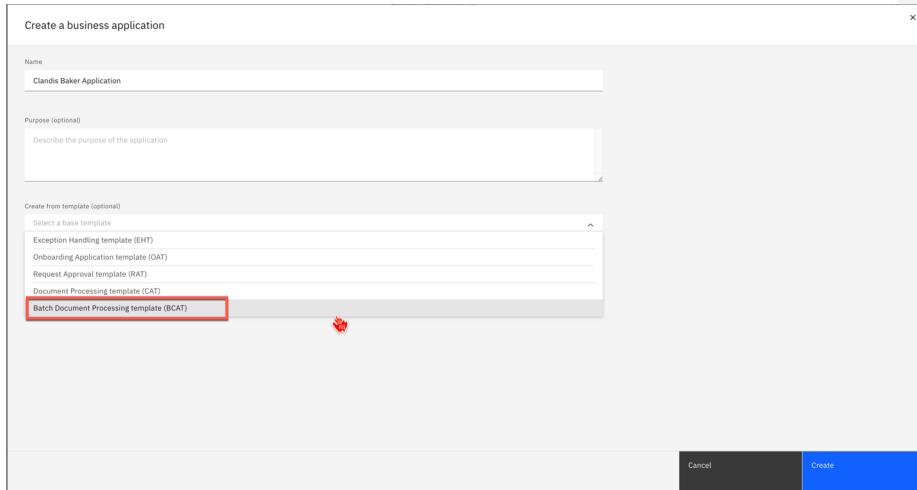
\_2. From the **Create** drop down list, select Application

A screenshot of the IBM Cloud Pak Administration interface under 'Business applications'. The left sidebar shows 'Create' (selected), 'Import', 'Application' (highlighted in red), 'Template', and 'Toolkits'. The main area shows a message: 'There are no applications to display...yet.' with a link to 'Create'. Below this are three template cards: 'Request Approval template', 'Onboarding Application template', and 'Exception Handling template', each with a 'Last updated' date of 02/20/2023.

\_3. Select **Enter your <application name>** in the Name field.

A screenshot of the 'Create a business application' dialog box. The 'Name' field contains 'Clandis Baker Application' (highlighted with a red box). The 'Purpose (optional)' field is empty. Under 'Create from template (optional)', there is a dropdown menu showing 'Select a base template' and a list of templates: 'Exception Handling template (EHT)', 'Onboarding Application template (OAT)', 'Request Approval template (RAT)', 'Document Processing template (CAT)', and 'Batch Document Processing template (BCAT)'. At the bottom right are 'Cancel' and 'Create' buttons, with 'Create' being highlighted in blue.

- \_4. In the Create Form Template in drop down **select Batch Document Processing template (BCAT)**.



You could have selected the Document Processing Template if you only wanted to process a single document at a time, but in this lab, you will process several documents in a batch.

- \_5. Click **Next**

- \_6. You will be presented with the Create an application window. In the Select repository **pick DEVOS1**

## Automation Document Processing Lab

Create an application

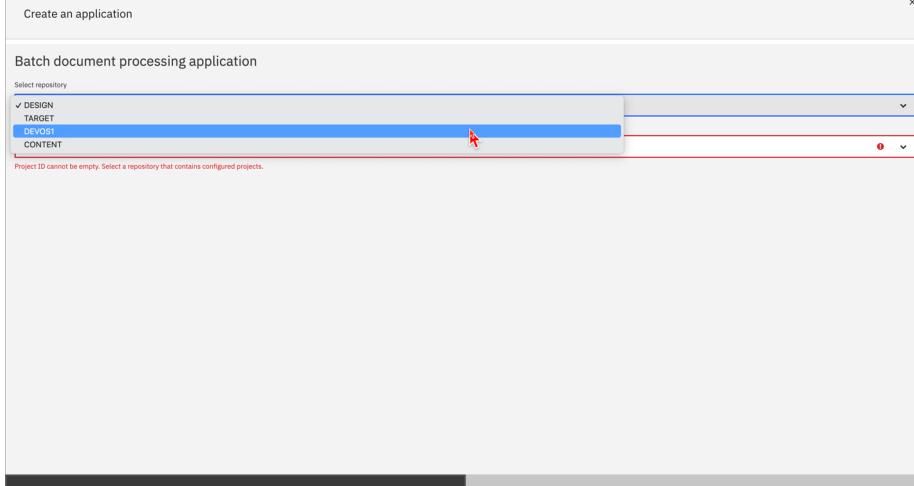
Batch document processing application

Select repository

✓ DESIGN  
TARGET  
**DEVOS1**  
CONTENT

Project ID cannot be empty. Select a repository that contains configured projects.

Back Create



\_7. In the Project ID drop down **pick <your project name>**.

Create an application

Batch document processing application

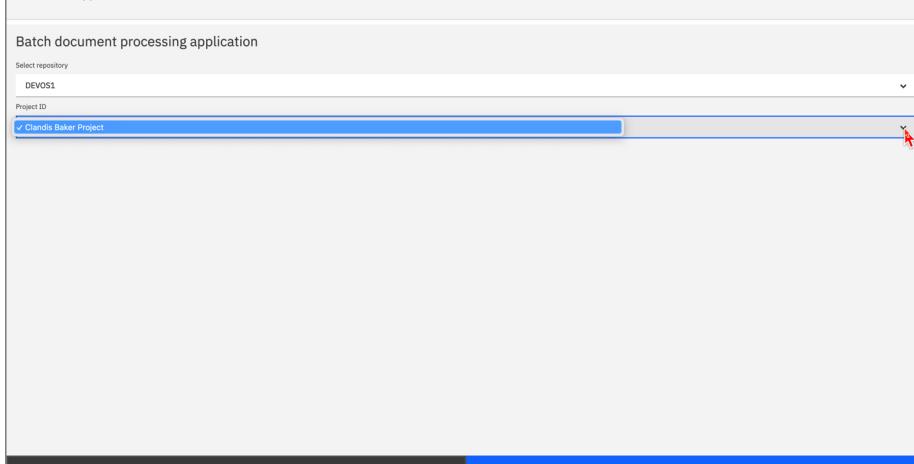
Select repository

DEVOS1

Project ID

Clandis Baker Project

Back Create

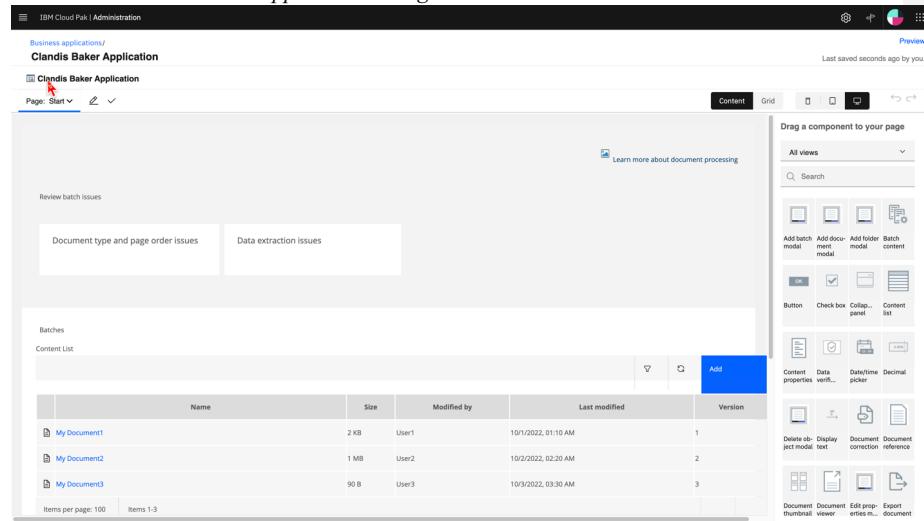


*Note it may take a minute or two before this update and you can see your project*

## Automation Document Processing Lab

### \_8. Click **Create**

You should now be in the *Application Designer*



The screenshot shows the IBM Cloud Pak Application Designer interface. At the top, there's a navigation bar with 'IBM Cloud Pak | Administration' and a 'Preview' button. Below the navigation is a breadcrumb path 'Business applications/Cländis Baker Application'. The main area displays a 'Content List' for 'Batches' with three items: 'My Document1', 'My Document2', and 'My Document3'. To the right of the content list is a sidebar titled 'Drag a component to your page' containing various UI component icons like 'Add batch modal', 'Content list', 'Content properties', etc. At the bottom left of the content list, there are pagination controls: 'Items per page: 100' and 'Items 1-3'.

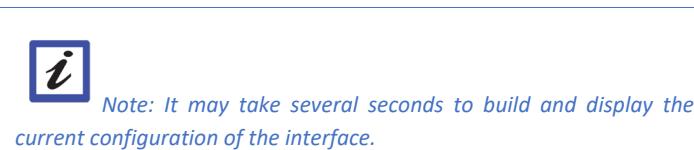


Batch Document Processing template (BCAT) has all the necessary pages and configuration to start using the application. Using this designer user interface, you have the option to further customize the application, such as its page design or actions, to fit your requirements.

### \_9. Click on **Business applications** breadcrumb at the top.

## Automation Document Processing Lab

The screenshot shows the IBM Cloud Pak Administration interface. In the top navigation bar, 'Business applications' is highlighted. The main content area displays the 'Clandis Baker Application'. On the left, there's a sidebar titled 'Drag a component to your page' containing categories like 'All views', 'Search', and 'Content properties'. The main content area has sections for 'Review batch issues' (Document type and page order issues, Data extraction issues) and 'Batches' (Content List). A table lists three documents: My Document1, My Document2, and My Document3, with columns for Name, Size, Modified by, Last modified, and Version. At the bottom, there are buttons for 'Items per page: 100' and 'Items 1-3'.

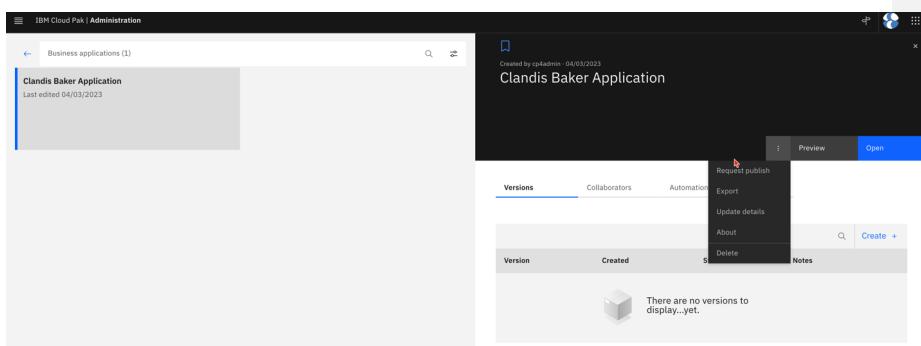


- \_10. From this screen you can also select preview to see the pre-configured interface. The open takes you the application where you can modify the look and feel. If you hover anywhere in the box it will turn grey and **Click**.

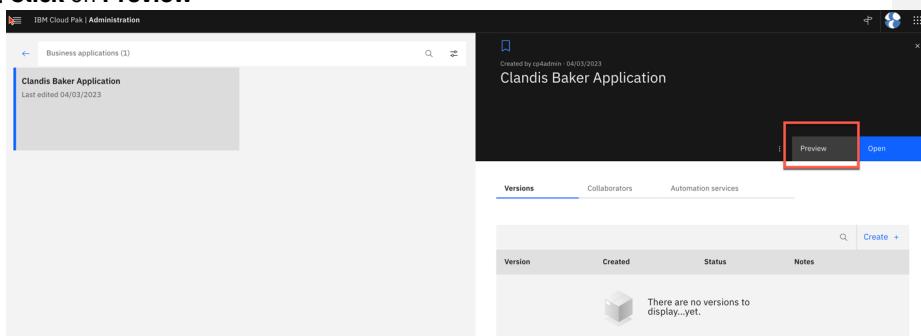
The screenshot shows the 'Business applications' list in the IBM Cloud Pak Administration interface. The 'Clandis Baker Application' is selected, and its details are shown in a modal window. The modal includes a 'Preview' button, which is highlighted with a blue border. Other buttons in the modal include 'Open' and 'Edit'. Below the modal, there are buttons for 'Create', 'Import', 'Applications', 'Templates', and 'Toolkits'.

- \_11. From this screen if you **click** on the **3 dots** you can save versions, export application, or delete the application. Just wanted to show this for future reference.

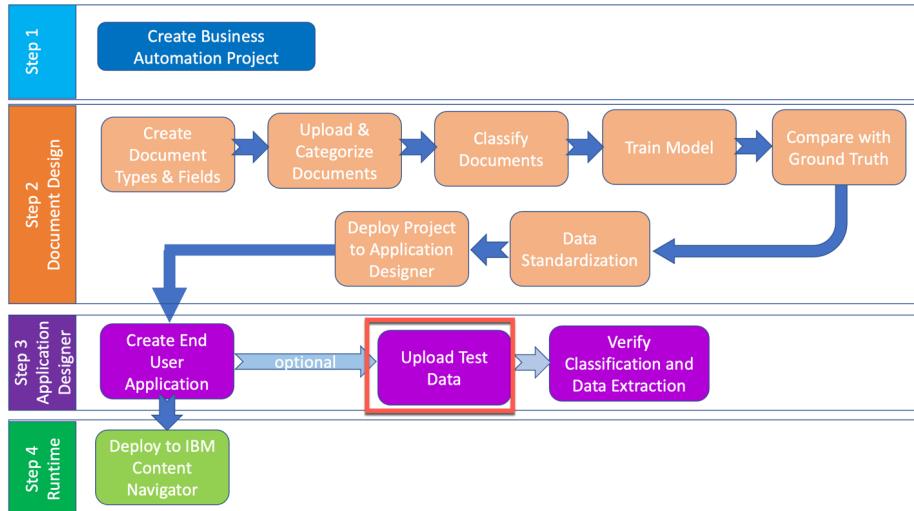
## Automation Document Processing Lab



### \_12. Click on Preview



## 11.2 Upload documents for processing



- \_1. You should be in the default application user interface for ADP it opens a new tab in your browser.

The screenshot shows the user interface for the Claudio Baker Application. At the top, there's a navigation bar with tabs like 'Cloud Pak for Integration', 'Administration', and 'Claudio Baker Application'. Below the navigation, there are two sections: 'Review batch issues' and 'Batches'.

**Review batch issues:**

- Document type and page order issues: 0 batches
- Data extraction issues: 0 batches

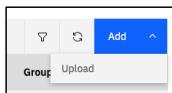
**Batches:**

Name	Files	Priority	Status	Added on	Added by	Group	Location
(empty)	(empty)	(empty)	(empty)	(empty)	(empty)	(empty)	(empty)

No items found.

There are two key screens you will work with: “*Document type and page order issues*” and the “*Data extraction issues*”. First, we need to upload some test documents and have them processed.

- \_2. Click on Add, then Upload.



- \_3. Enter a **name** for your batch in the Display Name field and set the **Priority to High** as seen in the image below.

A screenshot of a 'Upload new batch' form. It includes fields for 'Display Name' (containing 'Batch 1'), 'Description', and a 'Priority' dropdown menu set to 'High'. There is also a link to 'Upload'.

- \_4. Click **Select files**.

Navigate to the samples folder previously downloaded from [Section 2](#) and use the *Group 3 - Runtime Demo Set* folder documents. Select all the files in the folder.

- \_5. Click **Open**

You will see a window that will give the operator a chance to manually classify the documents before they are ingested. By clicking on one of the files you will be presented with an option to manually classify the documents. In the example below would be how to manually classify a document.

A screenshot of an 'Add Files' dialog. It shows a table with 1 item selected. The columns are 'File Name' and 'Document Type'. The first row has a checked checkbox next to 'B\_PO\_5.pdf' and 'Auto Classify' under 'Document Type'. Below the table are 'Classify' and 'Auto Classify' buttons, followed by a 'Deselect' button. At the bottom are 'Cancel' and 'Add' buttons. A note at the bottom states: 'We are not going to do this but instead let ADP auto classify them.'

## Automation Document Processing Lab

### Add Files

To manually specify document type, first select the files in the table. Use the classify option, to assign the document type for selected file(s). If a file is not manually classified, the system will auto-classify it.

The screenshot shows a table titled 'Filter List' with columns for 'File Name' and 'Document Type'. There are five rows, each with a checkbox in the first column. The 'Document Type' column contains the value 'Auto Classify' for all rows. The last row is 'DE\_FW2\_4001\_0001S.pdf'. At the bottom of the table are 'Cancel' and 'Add' buttons.

	File Name	Document Type
<input type="checkbox"/>	B_PO_5.pdf	Auto Classify
<input type="checkbox"/>	DE_FW2_1000_0001F.pdf	Auto Classify
<input type="checkbox"/>	DE_FW2_4000_0011F.pdf	Auto Classify
<input type="checkbox"/>	DE_FW2_4001_0001S.pdf	Auto Classify
<input type="checkbox"/>	DE_FW2_4001_0010F.pdf	Auto Classify

Cancel Add

### \_6. Click on the **Add** button.

The screenshot shows the 'Review batch issues' section with two tabs: 'Document type and page order issues' (0 batches) and 'Data extraction issues' (0 batches). Below this is a 'Batches' table with one item named 'Batch01'. A progress bar at the bottom of the table indicates '3 of 5 files processed'. The table includes columns for Name, Files, Priority, Status, Added on, Added by, Group, and Location. The 'Status' column shows a blue progress bar. At the bottom left is 'Items per page: 100' and '1-1 of 1 items'. At the top right is a 'Learn more about document processing' link.

Name	Files	Priority	Status	Added on	Added by	Group	Location
Batch01	5	High	<div style="width: 60%;">3 of 5 files processed</div>	02/23/2023, 10:49 AM	cp4admin		

A progress bar will be displayed indicating when all documents have been uploaded.

### \_7. Click the 3 dots at the end of the line.

The screenshot shows the 'Review batch issues' section with the same tabs and batches as the previous screenshot. The 'Batches' table now has a red arrow pointing to the three-dot menu icon in the 'Location' column of the 'Batch01' row. The rest of the interface is identical to the previous screenshot.

### \_8. Click Submit

In the screen shot below, you see we have a document issues (status) and we now have 1 batch in the “Document type and page order issue” tile.

Review batch issues

Document type and page order issues 1 batches

Data extraction issues 0 batches

Batches

Name	Files	Priority	Status	Added on	Added by	Group	Location
Batch01	5	High	Document issues	03/27/2023, 01:45 PM	cpadmin		

Items per page: 100 1-1 of 1 items

### 11.3 Correct any classification errors.

\_1. Click on the **Document type and page order issues** tile to open the batch.

Batch Document Processing Application /

Document type and page order issues

Name	Priority	Status	Added on	Added by	Group	Location
Batch 1	High	Document issues	01/13/2021, 08:44 am	CEAdmin		

Items per page: 100 1-1 of 1 items

\_2. Click on <your batch name> to open it.

You should now see all the documents you uploaded in your batch. The ones with issues will have a yellow checkmark for documents that have a low confidence document type and a red exclamation mark for documents it could not classify.

## Automation Document Processing Lab

Batch01

PURCHASE ORDER  
RUBE'S Meat Co.  
P.O. No. 7128  
DATE: 29 March 2020  
CUSTOMER ID: 409701

VENDOR:  
Chicken Run Ranch  
24 Quay Street  
Nelson Village NE23 8DD  
UK  
078-2054-8486

SHIPPING METHOD: AIR SHIPPING TERMS: C.I.F. DELIVERY DATE: 29 March 2020

QTY	ITEM #	DESCRIPTION	JOB	UNIT PRICE	LINE TOTAL
230 PCs	01	Whole Chicken		£1.50	£345.00
150 Packs	02	One Day Old Chick		£1.05	£157.50
					TOTAL £502.50

\_3. Most of the document types are correct but it looks like a PO got mixed into our batch so we can **Click** on the **Trash can** to delete it from the batch. And **select OK** to delete it.

Batch01

PURCHASE ORDER  
RUBE'S Meat Co.  
P.O. No. 7128  
DATE: 29 March 2020  
CUSTOMER ID: 409701

VENDOR:  
Chicken Run Ranch  
24 Quay Street  
Nelson Village NE23 8DD  
UK  
078-2054-8486

SHIPPING METHOD: AIR SHIPPING TERMS: C.I.F. DELIVERY DATE: 29 March 2020

QTY	ITEM #	DESCRIPTION	JOB	UNIT PRICE	LINE TOTAL
230 PCs	01	Whole Chicken		£1.50	£345.00
150 Packs	02	One Day Old Chick		£1.05	£157.50
					TOTAL £502.50

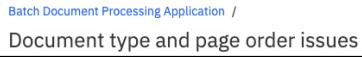
- \_4. Review all documents to ensure everything is correct. If the system no longer detects any issues, you should see a green checkmark near the top of the document list.



- \_5. **Click Save Changes** and then **Submit** to save your changes and have the batch processed.

The system will start reprocessing the documents now that they have been classified correctly.

- \_6. **Click** on the blue **Batch Document Processing Application link** at the top to return to the previous preview menu.



#### 11.4 Correct extraction issues

The following instructions are based on a pre-trained sample application. Not what you will see in your untrained application.



*Important Note: The project you are using for this has been configured but NOT run through the training (Deep Learning). So, the results will not reflect what they should be. IN A NORMAL SCENARIO, ON A CLUSTER WITH GPU AND DEEP LEARNING ENABLED, YOU WOULD HAVE TRAINED YOUR MODEL BEFORE DEPLOYING IT AND WOULD BENEFIT FROM HIGHER EXTRACTION RATES. the purpose of this lab is to teach you the tools but won't show you the trained results.*

It may take a few seconds for your batch to advance to the next step. If your batch needs further attention, you will see it appear in the Data extraction issues tile.

- \_1. **Click** on the **Data extraction issues** tile to open it.



## Automation Document Processing Lab

\_2. Click on <your Batch name> to open.



After opening we see all the documents that have been processed but one looks to have extraction issues.

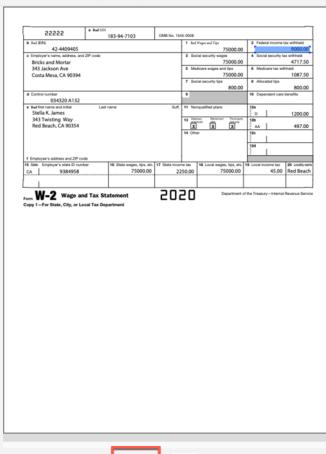
Batch Document Processing Application / Batches with data extraction issues /  
Batch01

Name	Issues	Status	Modified on	Modified by
BAD_FW2_1000_0003F.pdf	1	Data issues	03/04/2023	cp4admin
TR_FW2_1000_0003F.pdf		Issues reviewed	03/04/2023	cp4admin
TR_FW2_2000_0003F.pdf		Issues reviewed	03/04/2023	cp4admin
TR_FW2_4000_0002F.pdf		Issues reviewed	03/04/2023	cp4admin

Items per page: 100 1-4 of 4 items

\_3. Click on the bad document to open it. Zoom in a bit to get a better picture of the document.

BAD\_FW2\_1000\_0003F.pdf | Document type: Wage and Tax



Extracted data

All Fields

**Federal Income Tax Withheld**  
Federal Income Tax Withheld  
9000.00

**Employee Social Security Number \***  
(none)

**Employer Identification Number**  
(none)

**Employers Name and Address**  
Bricks and Mortar 343 Jackson Ave Costa Mesa, CA 90394

**Social Security Wages**  
75000.00

**Wages Tips Other Compensation**  
(none)

**Employee Name and Address \***  
Last name Suff. Stella K. James 343 Twisting Way Red Beach, CA 90354 | Employee's address and ZIP code

Take a moment to discover the image viewer features.

## Automation Document Processing Lab

### Image viewer features at top:

The screenshot shows a document processing interface with a PDF viewer at the top. The PDF is titled "Form W-2 Wage and Tax Statement" and includes fields like Employee Social Security Number (334-91-3066), Employer ID number (87-3849583), and Control number (EED640 FGH). Below the PDF, there is a toolbar with icons for back, forward, search, and other document operations.

- Rotate image.
- Visual effect adjustment
- Invert

### Image viewer features at bottom:

The screenshot shows a document processing interface with a PDF viewer at the top. The PDF is titled "W-2 Wage and Tax Statement" and includes fields like Federal Income Tax Withheld (9900.00), Social Security Wages (75000.00), and Wages Tips Other Compensation (none). Below the PDF, there is a toolbar with icons for back, forward, search, and other document operations.

- Page and thumbnail's view

- Fit to window.
- Zoom and Magnify

#### Field features

The screenshot shows a document processing application interface. On the left, there is a preview of an W-2 form for the year 2020. The form includes fields for employee information (Name: Stella K. James, Address: 343 Twisting Way Red Beach, CA 90354), employer information (Bricks and Mortar 343 Jackson Ave Costa Mesa, CA 90394), and tax details (Social Security Wages: 76000.00, Federal Income Tax Withheld: 9000.00). On the right, the 'Extracted data' panel is open. It has a dropdown menu at the top set to 'All Fields'. Below it, several fields are listed with their extracted values and confidence levels. Some fields have a red validation error icon next to them. The fields include:

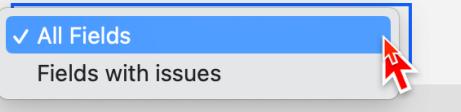
- Federal Income Tax Withheld: 9000.00 (confidence 100%)
- Employee Social Security Number: (none)
- Employer Identification Number: (none)
- Employers Name and Address: Bricks and Mortar 343 Jackson Ave Costa Mesa, CA 90394
- Social Security Wages: 76000.00 (confidence 100%)
- Wages Tips Other Compensation: (none)
- Employee Name and Address: Last name Suff, Stella K. James 343 Twisting Way Red Beach, CA 90354 / Employee's address and ZIP code

- Show all fields.
- Show fields with issues.

Also note that fields that do have issues have a notification icon next to them. For example, Wages Tips Other Compensation field picked up correctly but has a low confidence based on the extraction results.

#### \_4. Under Extracted data click on the drop down twisty.

##### Extracted data



#### \_5. Click on the ALL Fields.

This view shows all the fields that we defined earlier. Fields with an asterisk are mandatory fields.

**Change** the Extracted data back to **Fields with issues**.



The Employee Social Security Number is a mandatory field. For purposes of this lab it was changed to "Bad SSN". Since you did not make that phrase an alias ADP was not able to pick it up.

Commented [CB4]: Need to delete this. Add going through all docs

- \_6. Click on **Employee Social Security Number** and with your mouse **select** the SSN under “**Bad SSN**”.

The screenshot shows the ADP software interface. On the left, a W-2 form is displayed with various fields filled in. On the right, a 'Extracted data' dialog box is open, specifically focusing on the 'Fields with issues' section. Under the heading 'Employee Social Security Number \*', there is a single entry: '183-94-7103'. Below this entry, a message states 'Required value is missing.' There are also other sections visible in the dialog, such as 'Issue types' and a 'Learn more' link.

Also the Wages Tips Other Compensation did not have a correct alias defined. But since it was not a required field, you can continue to process.

- \_7.. Click on **Save Changes** box at the top.

Cancel

Save changes

Done and next

Done

(?) Learn more ...

- \_8. For the remaining fields there are no extraction issues that ADP picked up for mandatory fields. You may see some low confidence characters. If so, **Click** on Dismiss for each field with a yellow validation warning.

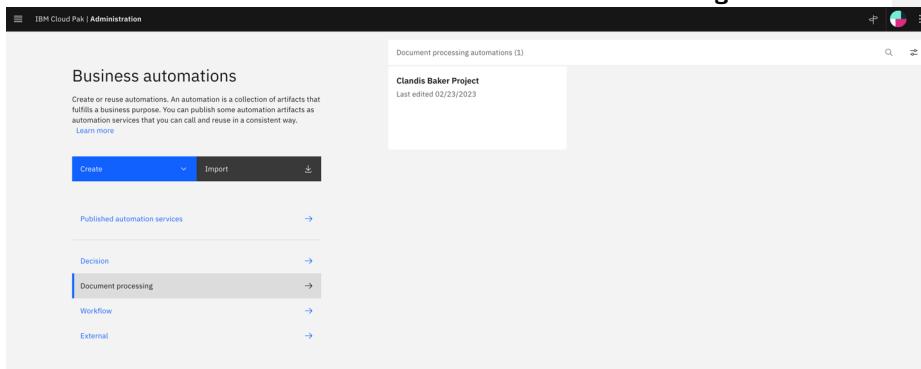
- \_9. **Click** on **Done and next**.

\_10. All documents have been processed **Click** on **Submit** at the top to complete the batch.

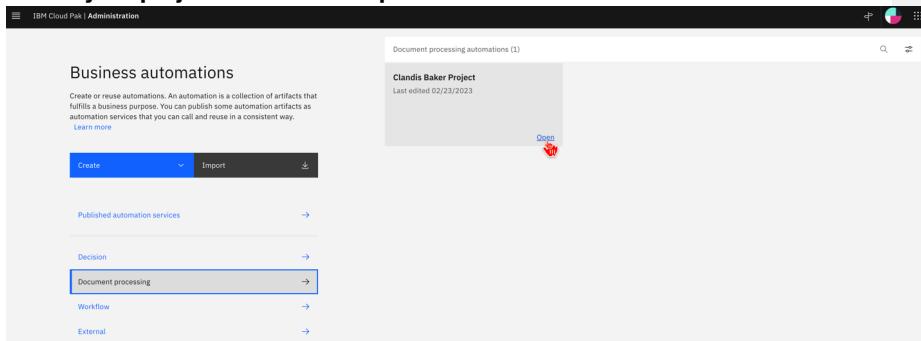
## 12 Optional Export/Import Project.

If you would like to save your project and perhaps use it later, you can do this lab.  
From the Business Automations

- \_1. From the Business Automations screen **select Document Processing**.

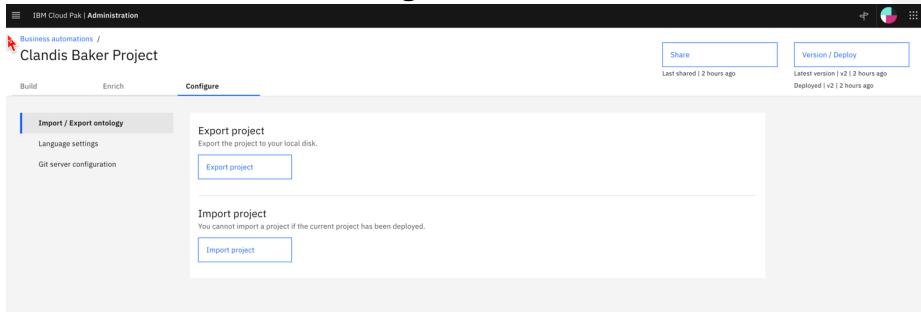


- \_2. Select <your project name> Click open



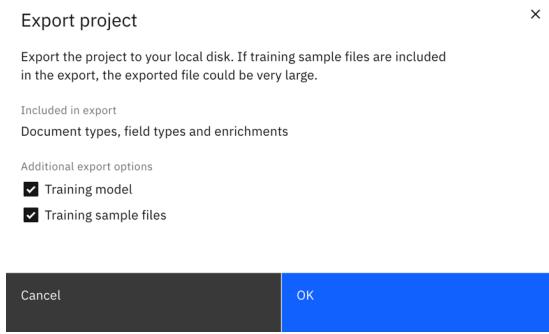
## Automation Document Processing Lab

\_3. From the Main screen **select the Configure tab**



\_4. **Select Export Project**

\_5. On Export Project window **check Training Module and Training Sample files**



\_6. **Click on OK**

\_7. A project-export-<date-time>.zip will be download via browser to local machine.

---

END OF LABS

---

## Appendix A - Troubleshooting

### TechZone Pending Status taking Long Time

Operator shows Pending status in a namespace – OLM know issue.

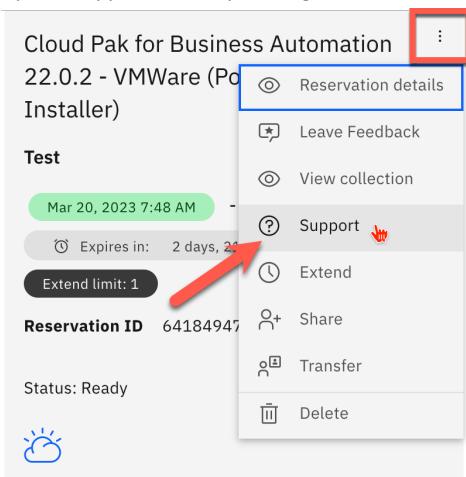
An operator fails to install and continuously shows Pending status.

For fix visit below link.

<https://www.ibm.com/docs/en/cpfs?topic=ii-operator-shows-pending-status-in-namespace-olm-known-issue>

Other issue could be the deployment itself had an issue. Two things to do in this case.

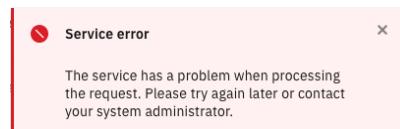
1. Open a support ticket by clicking on the 3 dots on the tile.



IBM Internal can also access support via SLAC Channel at #itz-techzone-support

2. Delete tile and try to deploy again.

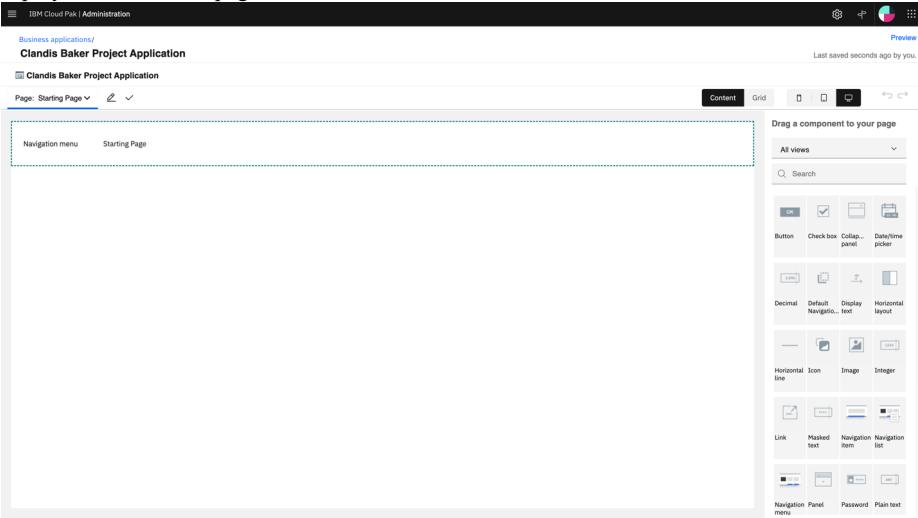
### Service Error



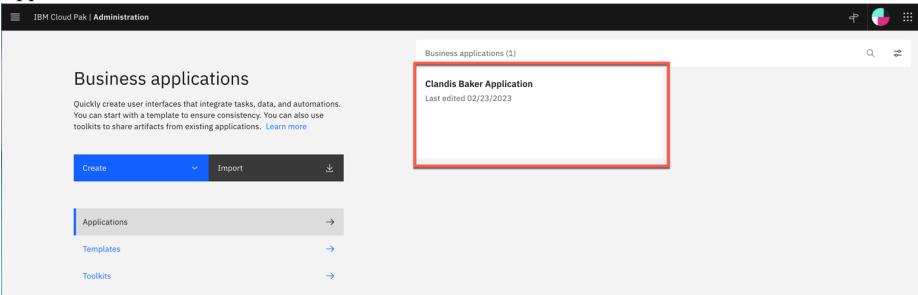
There was trouble communicating most of the time you can simply ignore and continue on.

## Application Blank

During creating of Business Application setup, sometimes on first time after project has been deployed. The Starter page is blank.

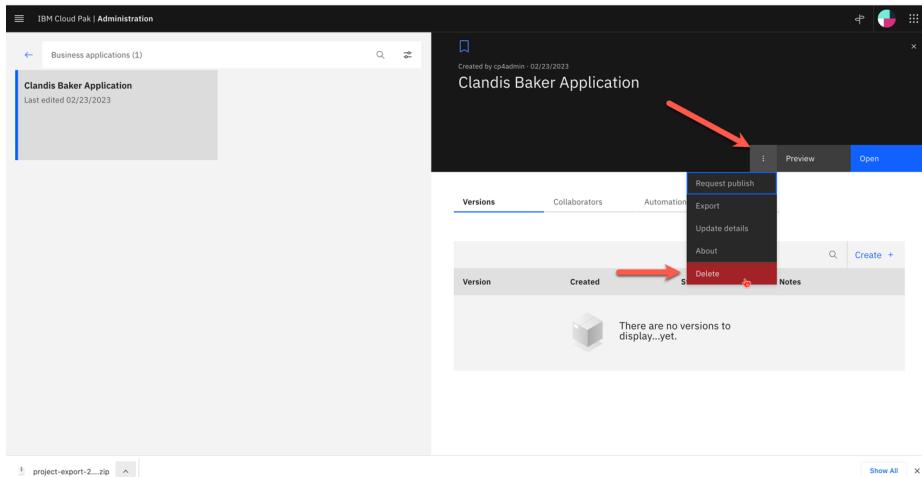


If this happens delete the application and try again. To delete the application, Click on the Application tile



Then Click on the 3 dots and Select Delete

## Automation Document Processing Lab



### Connection issue with Workstation to Cloud.

If issues with connection from workstation to cloud after it's been working. Reboot your workstation.

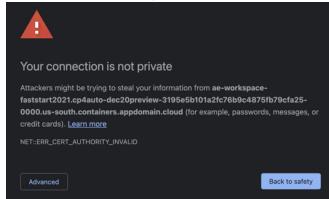
### Opening an Incognito Window

When you open a new incognito window, you will need to accept certificates before logging in to ADP. Customers shouldn't have this issue because they will have their own certificates instead of the self-signed certificates used in this environment.

In your incognito window, go to the following URLs located in this Box:

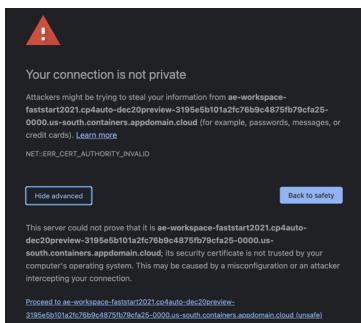
Open the Generate Security Tokens Box note and click all 3 of the links listed. This will reset the self-signed security certificates.

For each URL, your browser window will show a message like this:

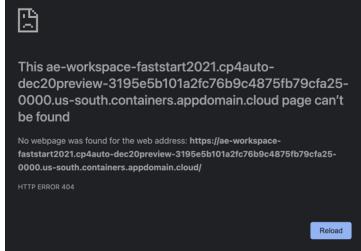


Click Advanced, and the browser window will look something like this:

## Automation Document Processing Lab



Click the “Proceed to...” link. You’ll see a message like this in your browser window:

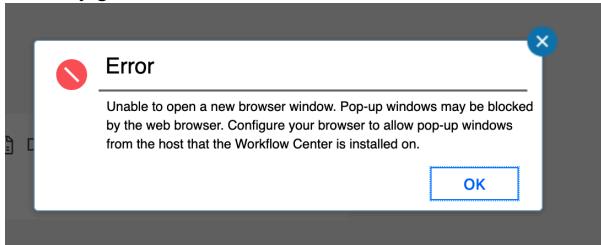


Ignore the error and proceed to the next link.

After doing this for each of the URLs above, log in to BAStudio

### Popup Blocked when trying to Preview Application.

You may get error like this:



You will need to grant access to pop up windows in your browser.

## **Appendix B - BAW & ADP Integration Sample**

For the End to End demo BAW was integrated with ADP. This link explains how to accomplish.  
<https://github.com/IBM/baw-adp-integration-sample>

## **Appendix C - Badge Information.**

Badge quiz page - <https://learn.ibm.com/user/policy.php>