

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 Set up WireGuard VPN .....	11
2.3 Open your IBM Cloud Environment .....	13
<b>3 Lab Overview.....</b>	<b>18</b>
3.1 How does ADP work? .....	18
<b>4 Create Document Processing Project .....</b>	<b>20</b>
4.1 Reviewing the interface.....	25
4.1.1 Build Tab .....	25
4.1.2 Enrich Tab .....	26
4.1.3 Configure Tab.....	27
<b>5 Configure a Wage and Tax document type. ....</b>	<b>30</b>
5.1 Create Wage and Tax document type.....	30
5.2 Create Field .....	32
5.3 Create the Employee Name Address field.....	34
5.4 Create Employee Social Security Number Field.....	35
<b>6 Document Types and Samples Overview.....</b>	<b>39</b>
6.1 Categorize documents. ....	40
<b>7 Train classification.....</b>	<b>47</b>
7.1 How do I improve my results? .....	51
7.1.1 Option 1 – Add more samples.....	51
7.1.2 Option 2 – review all uploaded samples.....	52
<b>8 Data extraction.....</b>	<b>53</b>
8.1 Correcting extracted values .....	56
8.2 Train extraction model .....	60
<b>9 Data standardization .....</b>	<b>61</b>
<b>10 Version and deploy your project .....</b>	<b>63</b>
<b>11 Application designer.....</b>	<b>65</b>
11.1 Create your Runtime Application .....	65
11.2 Upload documents for processing .....	70
11.3 Correct any classification errors.....	73
11.4 Correct extraction issues.....	75
<b>12 Optional Export/Import Project.....</b>	<b>80</b>
<b>Appendix A - Troubleshooting .....</b>	<b>82</b>
TechZone Pending Status taking Long Time .....	82
Can't find user/password in Daffy .....	82

Service Error .....	84
Application Blank .....	85
Connection issue with Workstation to Cloud.....	86
Opening an Incognito Window.....	86
Popup Blocked when trying to Preview Application. ....	87
<b>Appendix B - BAW &amp; ADP Integration Sample.....</b>	<b>88</b>
<a href="https://github.com/IBM/baw-adp-integration-sample">https://github.com/IBM/baw-adp-integration-sample</a> .....	88
<b>Appendix C - Badge Information. .....</b>	<b>89</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.

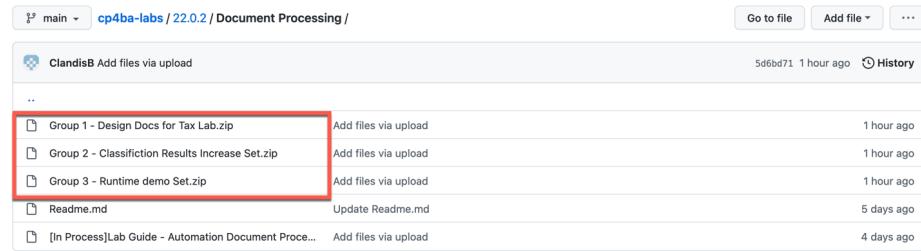
You can extract text, check boxes, forms, tables, barcodes, signature detection and even free text. With no or low code options, you can create an application that processes documents, extracts data, flags issues, and stores your documents and data. And the data enrichment capabilities ensure that the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

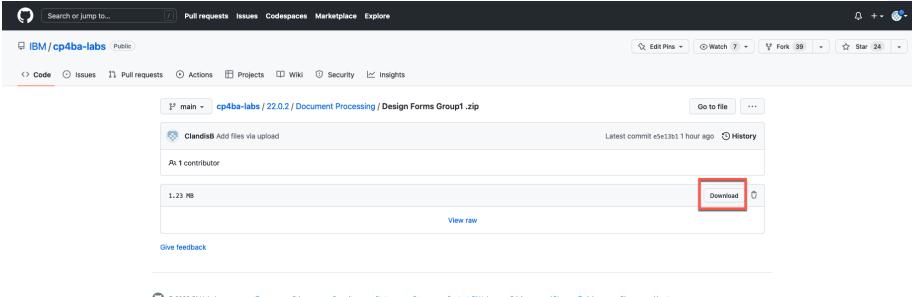
<https://github.com/IBM/cp4ba-labs/tree/main/22.0.2/Document%20Processing>



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.

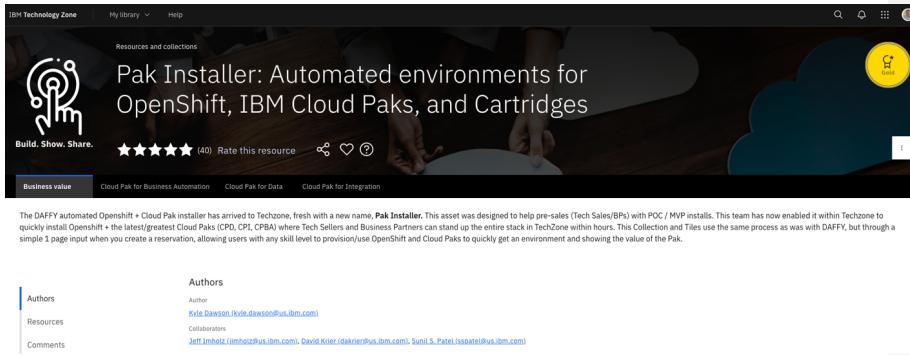
## Automation Document Processing Lab

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/>

### 2.1.1 Reserve Environment

- \_1. Navigate to <https://techzone.ibm.com/collection/63457fcba311ed0018ca2442>



The screenshot shows a collection page titled "Pak Installer: Automated environments for OpenShift, IBM Cloud Paks, and Cartridges". The page includes a rating of 4.5 stars from 40 reviews, a "Rate this resource" button, and three tabs: "Business value", "Cloud Pak for Business Automation" (which is selected), "Cloud Pak for Data", and "Cloud Pak for Integration". Below the tabs, there is a note about the DAFFY automated Openshift + Cloud Pak installer arriving in TechZone. The "Authors" section lists Kyle Dawson (kyle.dawson@us.ibm.com) as the author and Jeff Imboz (jimboz@us.ibm.com), David Krier (dakrier@us.ibm.com), Sunil S. Patel (sspatel@us.ibm.com) as collaborators.

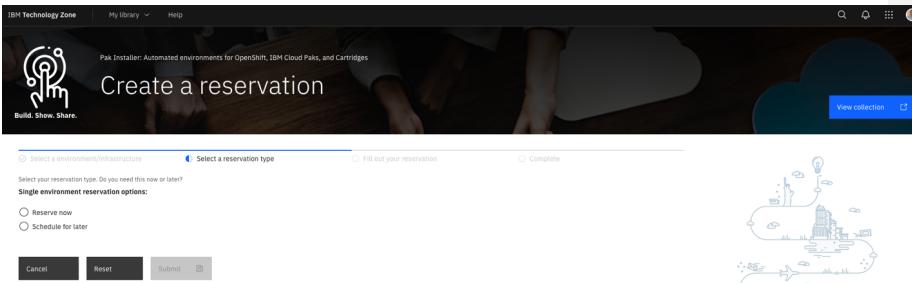


*Note: This environment is built with Daffy by Kyle Dawson with the latest releases. This environment can also be used at a customer site with same tool and framework of Daffy.*

- \_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

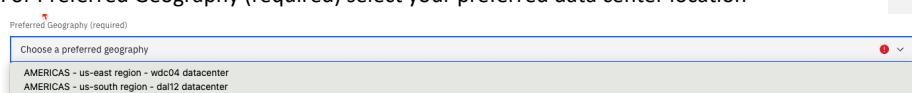
*Standard proof of concept*; Need an environment for a standard product use case.

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

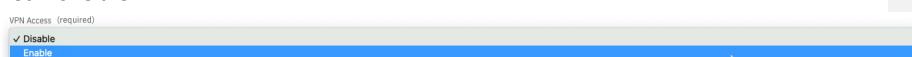
*Testing*: Need to test a specific function, configuration, or customization.

- \_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

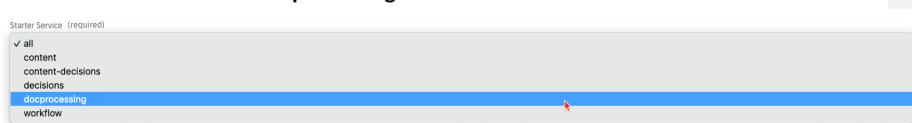


*Make sure to pick enable otherwise you'll have to start all over with deployment.*

- \_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



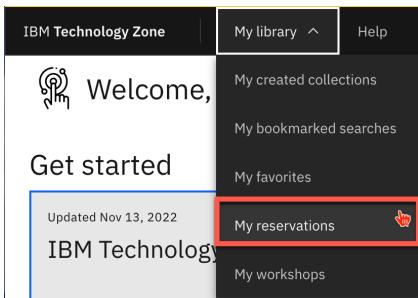
Upon receiving the Your environment is ready email, please allow up to 1 hour for the start-up services to fully complete. If after receiving email and a few hours have passed and your environment is not up, check [Appendix A – Troubleshooting](#) for possible fix.

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.

\_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**



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

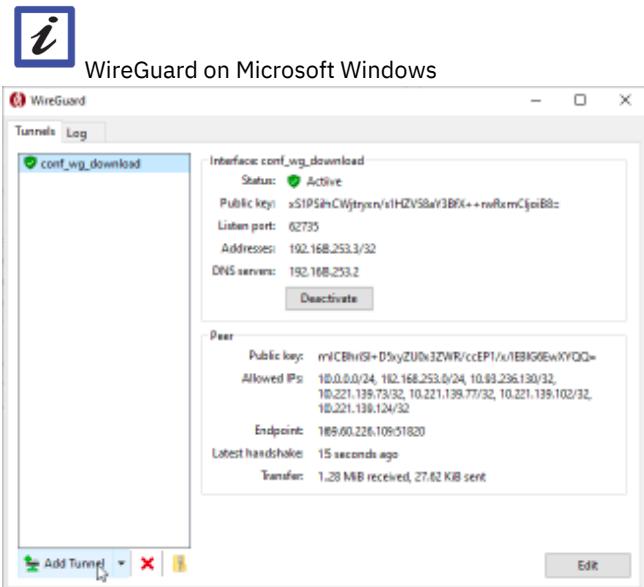
## 2.2 Set up WireGuard VPN

\_5. Open your reservation tile and scroll to bottom.

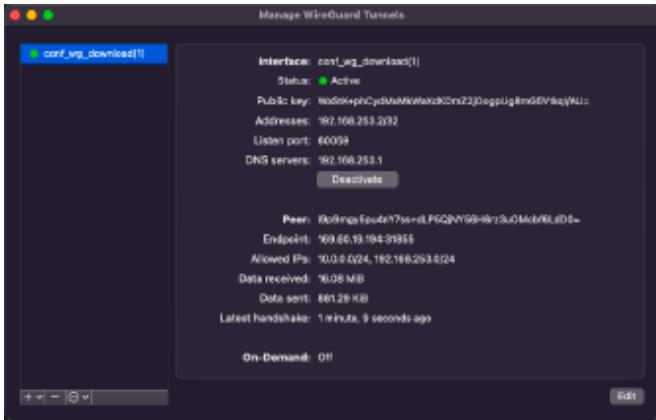
- \_6. Click Download WireGuard VPN config button to download conf\_wg\_download.conf to your local workstation

Download Wireguard VPN config 

- \_7. On your local workstation, install WireGuard by accessing <https://www.wireguard.com/install/>  
\_8. Launch WireGuard  
\_9. Click Add Tunnel and load the **conf\_wg\_download.conf** file.



## Automation Document Processing Lab



### 2.3 Open your IBM Cloud Environment

- \_1. Back on your reservation screen **Click on Open your IBM Cloud environment**

IBM Technology Zone My library Help

My reservations / Collection

Cloud Pak for Business Automation 22.0.1/22.0.2 - VMWare (Powered by Pak Installer)

Dates: Feb 26, 2022 7:40 AM Feb 27, 2022 7:40 AM Express in: 2 days, 22 hours, 41 minutes Extend limit: 2 Status: Ready

Desktop

Purpose: Open your IBM Cloud environment

Notes

Environment

Shared Reservation

For full desktop access, connect to: <https://remote.cloud.techzone.ibm.com/guacamole/#/username=gymuser-a09hjws&password=zHCBrb0y>

- \_2. Let's get the username and password created by DAFFY. **Expand OCP Gym under All Connections.**



If you're not familiar with Apache's Guacamole project, it is a clientless remote desktop gateway that allows you to access your desktop from any web browser.

## ALL CONNECTIONS

- ⊖ OCP Gym
- ⊖ gym-06000180cg-ar09hjws
  - \_ CloudPak Information
  - Remote Red Hat Linux Desktop
  - \_ SSH Command Line

### \_3. Select CloudPak Information

## ALL CONNECTIONS

- ⊖ OCP Gym
- ⊖ gym-06000180cg-ar09hjws
  - \_ CloudPak Information
  - Remote Red Hat Linux Desktop
  - \_ SSH Command Line

### \_4. This will open Daffy Options window. Enter 2 for Services

```
#####
#          Daffy Options          #
#####
Please use this tool and select what option you would like to retrieve more info on.
With this menu you can get your OpenShift Console URL, id/passwords and status.
You can also get your CP4BA Console URL, id/passwords and status info.

1) OpenShift
2) Services
3) ExitMenu
#? 2
```

5. Enter 1 for Console information

```
#####
#          Daffy Options          #
#####
Please use this tool and select what option you would like to retrieve more info on.
With this menu you can get your OpenShift Console URL, id/passwords and status.
You can also get your CP4BA Console URL, id/passwords and status info.

1) OpenShift
2) Services
3) ExitMenu
#? 2
CP4BA Services Menu:
1) Console
2) Status
3) Back
#? 1
```

6. Locate Username and Password and copy and paste these to notepad. You will need to login into your environment.

```
Validate CP4BA version info
#####
✓ PASSED  Valid version CPBA_VERSION=22.0.2
✓ PASSED  Valid IFIX CP4BA_IFIX=IF002

CP4BA Service Status
#####
Daffy Version           : v2023-03-09
Bastion OS              : rhel - 8.7
Platform Install Type   : vsphere-ipi
OpenShift Cluster Name  : ocpinstall
OpenShift Version        : 4.10.36
CP4BA Version           : 22.0.2 IF002
Project/Namespace       : cp4ba-starter
Zen Version              : 4.8.1
Message 1                : Running reconciliation
Message 2                : Prerequisites execution done.
Message 3                :
Message 4                :
Deployment Service      : Starter docprocessing
Config Map Dump          : /data/daffy/log/ocpininstall/cp4ba/icp4adeploy-cp4ba-access-info

Console Automation Document Processing
#####
Cloud Pak Business 3     : https://cp4ba-starter.apps.ocpininstall.gym.lan
Cloud Pak Admin Username  : cp4admin
Cloud Pak Admin Password  : Tm1WRtxkUI1bv2drooMF

#####
End Time: Mon Mar 20 11:43:53 EDT 2023
CP4BA Service Completed in 10 seconds
#####

CP4BA Services Menu:
Please select 1,2 or 3
#####
1) Console
```



*Note: Controls for copy and paste in guacamole.*

For Mac users:

**CONTROL\_OPTION\_SHIFT**

For Windows users:

**CTRL\_ALT\_SHIFT**



If your screen shows FAIL then it's not ready just yet and wait a bit longer.

```
# login https://api.ocpinstall.gym.lan:6443 -u ocpadmin -p ***** --insecure-skip-tls-verify
admin user - ocpadmin

Validate OCP Access
#####
✓ PASSED Access to cluster via oc command

Validate CP4BA version info
#####
✓ PASSED Valid version CPBA VERSION=22.0.2
✓ PASSED Valid IFIX CP4BA_IFIX=IF002

CP4BA Service Status
#####
Daffy Version : v2023-03-09
Bastion OS : rhel - 8.7
Platform Install Type : vsphere-ipi
OpenShift Cluster Name : ocpinstall
OpenShift Version : 4.10.36
CP4BA Version : 22.0.2 IF002
Project/Namespace : cp4ba-starter
Zen Version : 4.8.1
Message 1 : Running reconciliation
Message 2 : Prerequisites execution done.
Message 3 : FAIL - prerequisites Deployment failed ←
Message 4 :
Deployment Service : Starter docprocessing
Config Map Dump : /data/daffy/log/ocpinstall/cp4ba/icp4adeploy-cp4ba-access-info.yaml

Console Automation Document Processing
#####
Cloud Pak Dashboard : 
```

- \_7. Back on your Reservation tile **copy or click the link Cloud Pak Dashboard URL** and paste your favorite browser.

Cloud Pak Dashboard URL

<https://cpd-cp4ba-starter.apps.ocpinstall.gym.lan>

You may get a Your connection is not private, if click advance then click Proceed/Accept the Risk and Continue. This may occur twice.



### Warning: Potential Security Risk Ahead

Firefox detected a potential security threat and did not continue to **cpd-cp4ba-starter.apps.ocpinstall.gym.lan**. If you visit this site, attackers could try to steal information like your passwords, emails, or credit card details.

#### What can you do about it?

The issue is most likely with the website, and there is nothing you can do to resolve it.

If you are on a corporate network or using antivirus software, you can reach out to the support teams for assistance. You can also notify the website's administrator about the problem.

[Learn more...](#)

[Go Back \(Recommended\)](#)

[Advanced...](#)

- \_8. Login with <**Cloud Pak Admin Username /Cloud Pak Admin Password**> provided in step 6 above.

### 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.

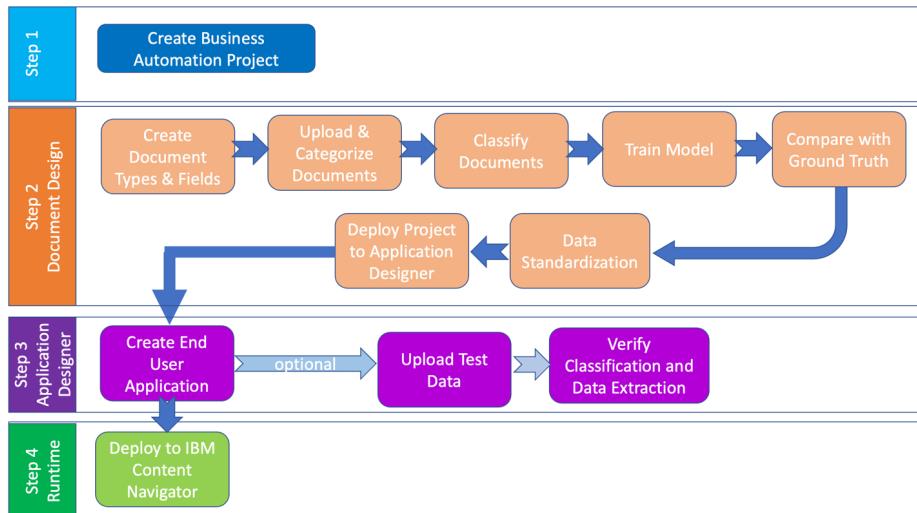
When you deploy your document processing application, you connect it to a content repository that manages the document types and the extracted data for each document. The solution is fully integrated with IBM FileNet® Content Manager, simplifying document and data storage by applying your existing filing 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

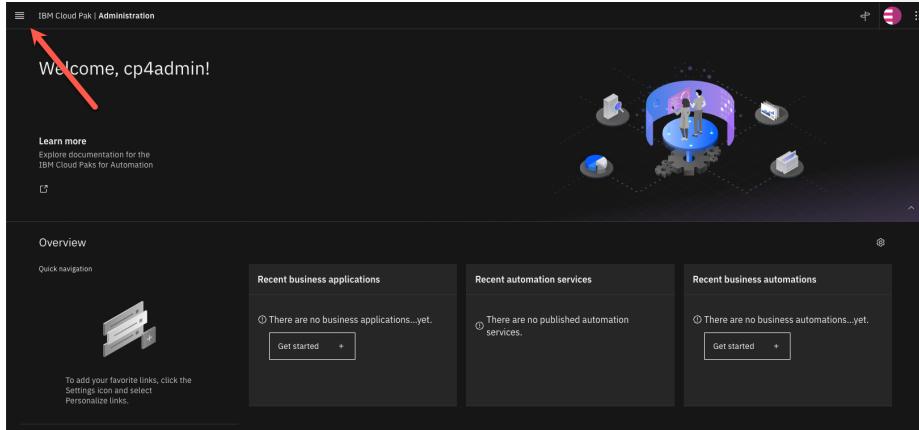
Step 1

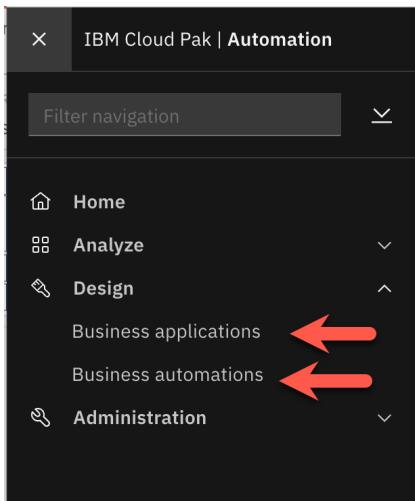
Create Business Automation Project

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.

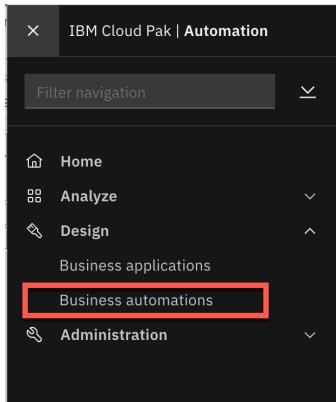
Within **Business Applications** you can quickly create user interfaces that integrate tasks, data, and automations. You can start with a template to ensure consistency. You can also use toolkits to share artifacts from existing applications.

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

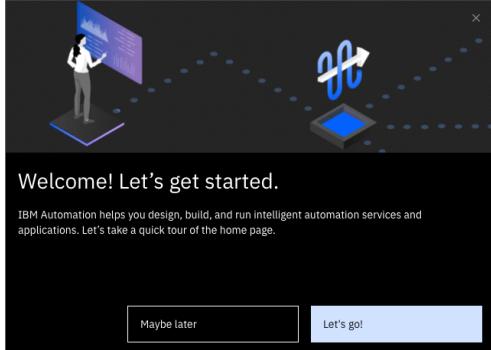
## Automation Document Processing Lab



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



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



Then following screen appears.

The screenshot shows the 'Business automations' page. At the top, there's a header bar with the IBM Cloud Pak logo and the word 'Automation'. Below the header, the title 'Business automations' is displayed. A brief description follows: '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 present. Below the description, there are two main buttons: 'Create' (highlighted in blue) and 'Import'. Underneath these buttons is a section titled 'Published automation services' with a right-pointing arrow. Further down, there are four categories listed with arrows: 'Decision', 'Document processing', 'Workflow', and 'External'.

\_9. Click on the **Create** twisty and select **Document processing automations**.

## Business automations

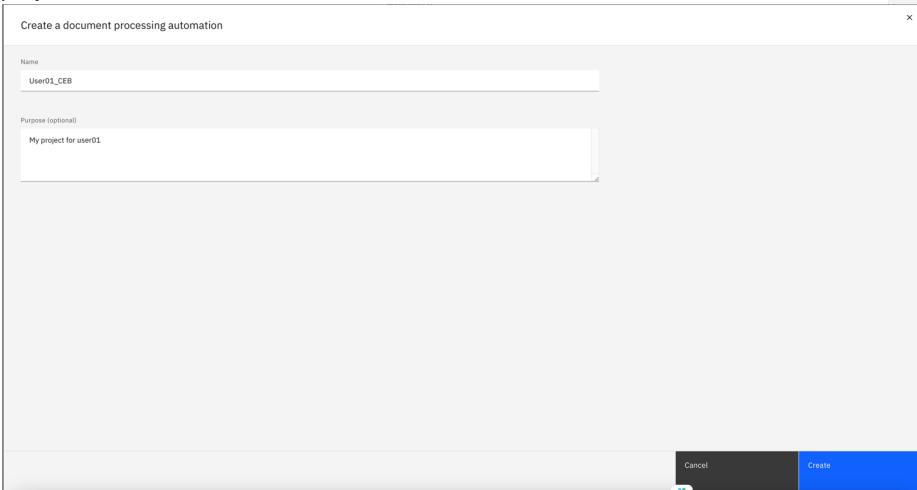
Create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way.

[Learn more](#)

This screenshot is similar to the one above, but the 'Create' button is highlighted with a blue background and a red arrow points to it from the left. A dropdown menu has opened, listing several options: 'Decision automations', 'Document processing automations' (which is highlighted with a red box), 'Workflow', and 'External'. Below this dropdown, the 'Published automation services' section and the other category rows ('Document processing', 'Workflow', 'External') are visible.

## Automation Document Processing Lab

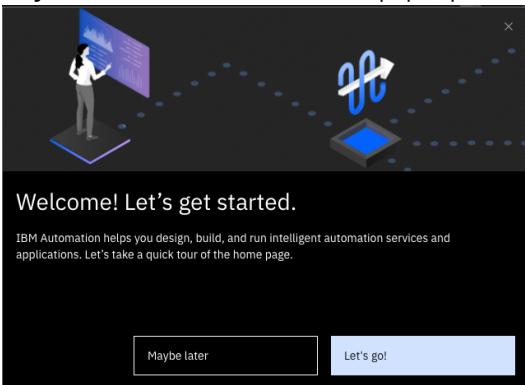
- \_10. In the Create a document processing automation window **enter a name** for the project.



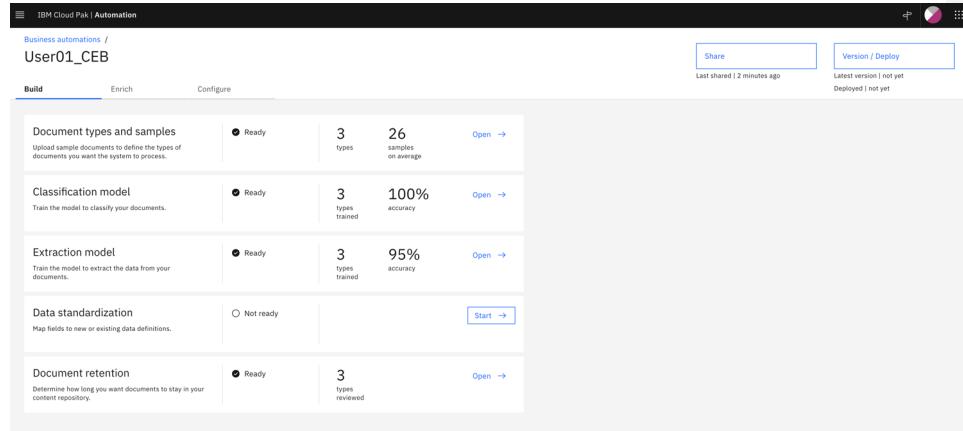
- \_11. Click on **Create** in the lower right-hand corner.



You may see the *Welcome Let's get started* throughout the lab simply **click Maybe later** whenever this window pops up.



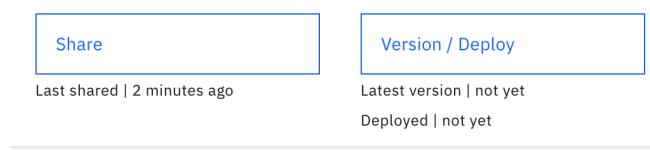
## 4.1 Reviewing the interface.



The screenshot shows the IBM Cloud Pak | Automation interface. At the top, there's a navigation bar with 'IBM Cloud Pak' and 'Automation'. Below it, a breadcrumb trail says 'Business automations / User01\_CEB'. On the left, there are three tabs: 'Build' (selected), 'Enrich', and 'Configure'. The main area is titled 'Document types and samples' with a sub-instruction 'Upload sample documents to define the types of documents you want the system to process.' It shows a status 'Ready' with '3 types' and '26 samples on average', with a 'Open' button. Below it are sections for 'Classification model' (status Ready, 3 types trained, 100% accuracy), 'Extraction model' (status Ready, 3 types trained, 95% accuracy), 'Data standardization' (status Not ready, with a 'Start' button), and 'Document retention' (status Ready, 3 types reviewed, with an 'Open' button). In the top right corner, there are 'Share' and 'Version / Deploy' buttons. Below them, status information says 'Last shared | 2 minutes 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.



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 we will be spending most of our 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 we 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.

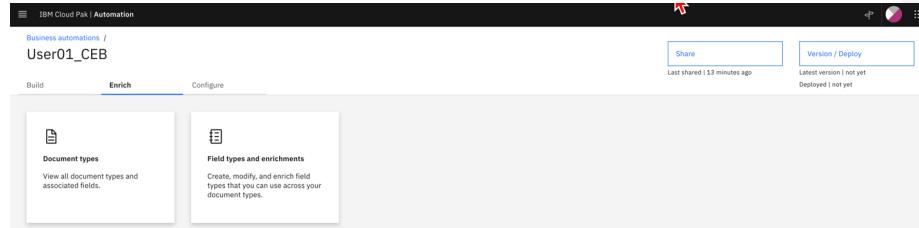
**Data Standardization:** This allows further refinement of the extracted information. For example, we want to standardize all dates to be formatted as YYYY/MM/DD. Having a standardized data format will help with any subsequent automation process.

**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.

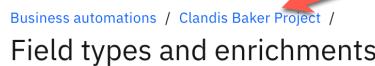


##### \_2. Click on FIELD TYPES AND ENRICHMENTS to begin. In this tile, you will see some of the pre-configured fields in the SYSTEM LIBRARY (sys). Customers can use these fields in their document type field definitions as needed.

## Automation Document Processing Lab

The screenshot shows the 'Field types and enrichments' section of the IBM Cloud Pak | Automation interface. On the left, there's a sidebar with 'Field type libraries' expanded, showing 'sys' selected. Below it are sections for 'Natural language extractors' and 'All libraries'. A specific library, 'U01CEB', is also listed under 'NamedEntityRecognition'. The main content area has a header 'Field types' with a search bar. Below it is a table with columns 'Field type' and 'Value type'. The table lists various field types like 'Address block', 'Address information', 'Addressee', etc., each with its corresponding value type (e.g., String, Composite, Boolean). To the right of the table is a detailed view of the 'Address block' enrichment, which includes sections for 'General' (Display name: Address block, Symbolic name: AddressBlock), 'Validator' (Validators: 4), and 'Value format' (Text: Extractors: None, Formatters: 3, Converters: None). At the bottom of the main content area, there's a note: 'You can modify existing field types within your libraries or create new ones. If data is formatted differently between documents, you can use field enrichments to reformat the extracted values and ensure consistency.'

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



### 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

The screenshot shows the 'Configure' tab for the 'Clandis Baker Project'. The 'Import / Export ontology' section includes 'Export project' and 'Import project' buttons. The 'Language settings' section is currently selected, showing 'Extraction language' and 'Display name language' settings. The 'Git server configuration' section is also visible.

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 screenshot shows the 'Configure' tab for the 'Clandis Baker Project'. The 'Language settings' section is selected, displaying 'Extraction language' and 'Display name language' configurations. Under 'Extraction language', 'English' is checked as the default. Under 'Display name language', 'English (en) (default)' is selected. Buttons for 'Cancel' and 'Save' are at the bottom.

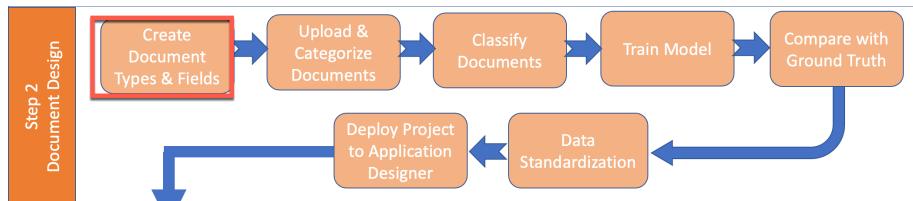
The Git server configuration is where you create a connection to the Git server for the first

## Automation Document Processing Lab

project that you create in Document Processing Designer. This setting applies to all subsequent projects that you create.

The screenshot shows the IBM Cloud Pak Administration interface for the 'Clandis Baker Project'. The top navigation bar includes 'IBM Cloud Pak | Administration', 'Business automations /', and a user icon. Below the navigation, there are tabs for 'Build', 'Enrich', and 'Configure', with 'Configure' being the active tab. On the left, a sidebar lists 'Import / Export ontology', 'Language settings', and 'Git server configuration', with 'Git server configuration' being the selected item. The main content area displays a form for configuring a Git server. It includes 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'). A 'Credentials' field contains the placeholder 'Enter a password or API key'. At the bottom of the form are 'Test' and 'Save' buttons. In the top right corner of the main content area, 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.

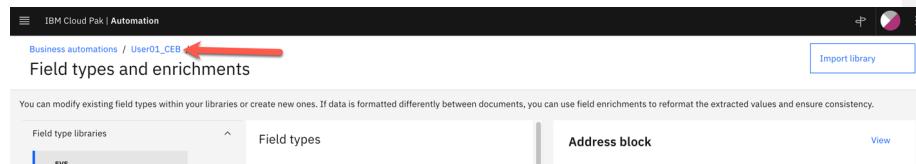


Before we use the guided configuration, you will configure some additional document types and fields used to extract data prior to uploading sample documents.

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 **<your project name>** in the breadcrumb trail to return to the start page. In the example below our project was called **<User01\_CEB>** if not already on the Project page



- \_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 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

The screenshot shows the 'Document types' section of the IBM Cloud Pak for Automation interface. On the left, there's a sidebar with 'User01-CEB' and 'Document types' sections, listing 'Bill of Lading', 'Invoice', 'Utility Bill', and 'Wage and Tax'. The main area is titled 'Wage and Tax' under the 'General' tab. It displays a message: 'You haven't added any fields yet.' with a link 'Click the Add field button to get started.' and a 'Add fields' button.

\_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

This screenshot is identical to the one above, showing the 'Wage and Tax' document type configuration. A red arrow points to the 'Add fields' button at the bottom of the central panel.

\_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 document type selected is 'Purchase Orders'. The 'General' tab is active. In the 'Display name' field, the value 'Ex. Employee's name, Le nom de l'employé' is entered, which is highlighted with a red border and has a red exclamation mark icon indicating it is a required field. The 'Description (optional)' field is empty. In the 'Symbolic name' field, 'Enter a name' is typed. The 'Field type' dropdown is set to 'sys:String'. Under 'Aliases', the values '2 Federal income tax withheld' and '2. Federal income tax' are listed, separated by a plus sign. There are also checkboxes for 'This field is required' and 'This field contains sensitive information'.

- 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 with the 'General' tab active. 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. The 'Value settings' tab is visible at the top.



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.



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.

- \_3. Click the **NEXT** button.
- \_4. Click **NEXT** again on the Field patterns screen. You will not be adding patterns in this lab. You should now be on the **VALUE SETTINGS** page. This is where you can set up validators, formatters, and converters. You will not be adding any value setting in this lab.
- \_5. Click **Create** your screen should look like this with your first field created.

Name	Type	Required	Sensitive
Federal Income Tax Withheld	Decimal	true	false

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:Address information**
- 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

The screenshot shows the 'Employee name and address' field configuration. The 'Display name' is 'Employee name and address'. The 'Aliases' section contains the field name and several suggested aliases: 'Employee's first name and initial Last name Suff', 'Employee's name, address, and ZIP code', 'Employee's name, address, and ZIP code', 'Employee Name & Address', and 'Employee's first name and initial'. The 'Field type' is 'sys:Address information' and 'Required' is checked.

- \_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 'Business automations / User01-CEB / Document Types / Employee Social Security Number' screen. At the top right are 'Cancel' and 'Next' buttons. Below them are three tabs: 'General' (selected), 'Field patterns', and 'Value settings'. The 'General' tab has sections for 'Display name' (Employee Social Security Number) and 'Symbolic name' (EmployeeSocialSecurityNumber). It also includes fields for 'Description (optional)' and 'Aliases'. Under 'Field type', 'sys:Numeric' is selected. There are checkboxes for 'This field is required' (checked) and 'This field contains sensitive information' (unchecked). The 'Aliases' section lists several entries separated by commas: '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'.

\_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

<b>Display Name</b>	<b>Description</b>	<b>Type</b>	<b>Mandatory</b>	<b>Aliases</b>
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>

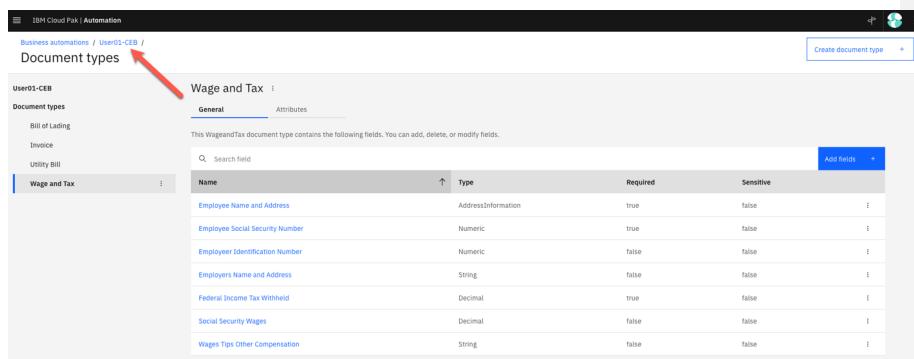
Reference for various field types:



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

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

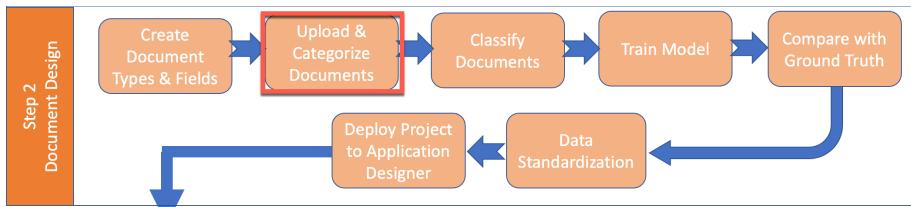
- \_6. Click on the <name of your project> in the breadcrumb link in the top left of your screen. In the following example the name of the project is <User01\_CEB>.



The screenshot shows the 'Document types' section of the IBM Cloud Pak for Automation interface. The breadcrumb navigation bar at the top displays 'IBM Cloud Pak | Automation' and 'Business automations / User01-CEB /'. Below this, the main title 'Document types' is followed by a sub-section title 'Wage and Tax'. There are two tabs: 'General' (selected) and 'Attributes'. A note below the title states: 'This WageandTax document type contains the following fields. You can add, delete, or modify fields.' A search bar and a 'Add fields' button are also present. On the left, a sidebar lists other document types: 'User01-CEB', 'Document types', 'Bill of Lading', 'Invoice', 'Utility Bill', and 'Wage and Tax' (which is currently selected). The main content area shows a table of fields:

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

Section	Status	Details
Document types and samples	Ready	4 types, 19 samples on average
Classification model	Ready	3 types trained, 100% accuracy
Extraction model	Ready	3 types trained, 95% accuracy
Data standardization	Not ready	
Document retention	Ready	4 types reviewed

## 6.1 Categorize documents.

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

- \_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 tabs for "Build" (which is selected), "Enrich", and "Configure". On the right, there are buttons for "Share" (Last shared | 2 days ago) and "Version / Deploy" (Latest version | not yet Deployed | not yet). The main area has a section titled "Document types and samples" with a sub-instruction: "Upload sample documents to define the types of documents you want the system to process.". This section is highlighted with a red box. Below it are four cards: "Classification model" (Ready, 3 types trained, 100% accuracy), "Extraction model" (Ready, 3 types trained, 97% accuracy), and "Data standardization" (Not ready). There is also a link to "Document types and samples" with a red arrow pointing to the "Document types and samples" section above.

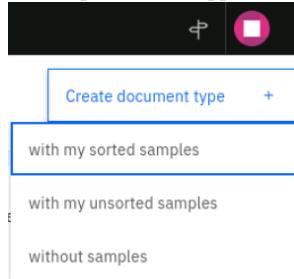
The CATEGORIZE feature analyzes each document and tries to find similarities between them. Based on these similarities, the system will divide the samples into categories for you to review. You can add documents or entire categories into either an existing document class or create new classes as needed.

Let's see what that looks like.

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



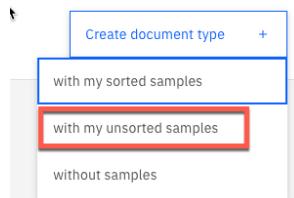
The drop down that appears:



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 either drag the folder to the window or 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\*](#)

Commented [CB1]: Show drag and drop and upload

*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.*

- \_5. Click on the **CATEGORIZE** button.

## Automation Document Processing Lab

The screenshot shows a web-based application titled 'IBM Cloud Pak | Automation'. The current page is 'Business automation / User interface / Document types and samples / Create document types'. At the top right are 'Cancel' and 'Categorize' buttons. Below the buttons 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.' There is a search bar labeled 'Search sample documents' and an 'Upload' button with a file icon. A list of 12 PDF files is displayed in a table:

Document name
Mortgage Agreement1.pdf
Mortgage Agreement2.pdf
Mortgage Agreement3.pdf
Mortgage Agreement4.pdf
Mortgage Agreement5.pdf
TR_FW2_1001_0000_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
UBILLCable_081_1_11.pdf
UBILLCable_082_1_11.pdf

At the bottom left are 'Items per page' and '20' dropdowns, and a page navigation bar showing '1 - 12 of 12 items'.



*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.

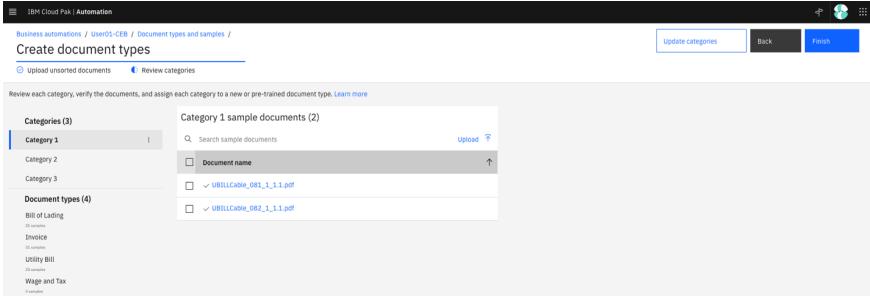
## Automation Document Processing Lab

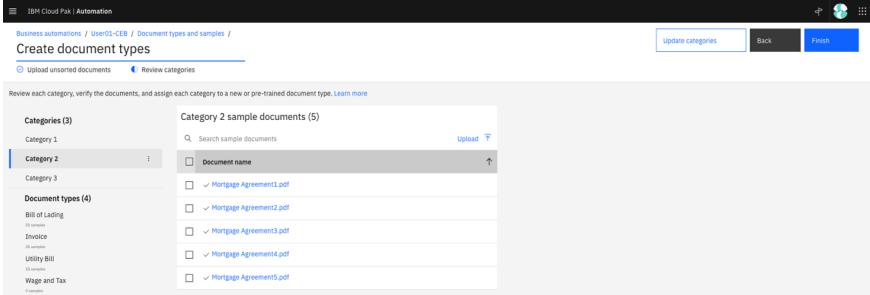
After a few seconds, the system will mark the documents with a status of ready as seen in the above image.

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

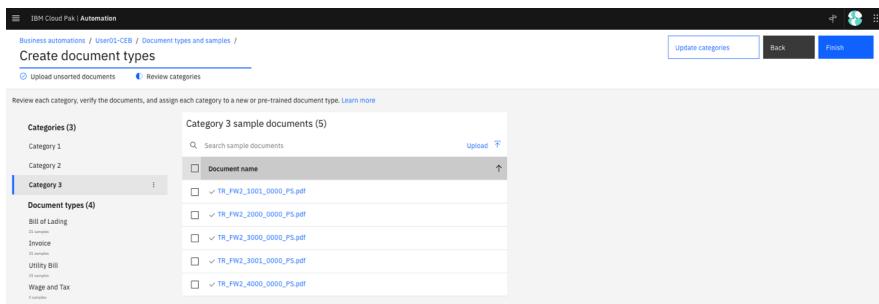
You can Click on any document to see a preview of it. This will help ensure the documents are correctly grouped.

**i** *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.*





## Automation Document Processing Lab

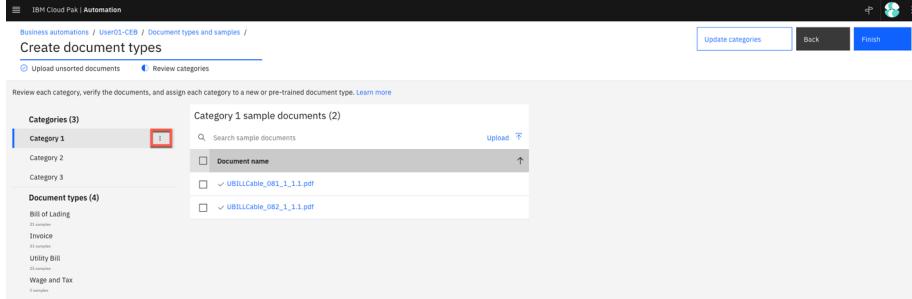


The screenshot shows the 'Create document types' page in the IBM Cloud Pak | Automation interface. The left sidebar lists categories: 'Categories (3)' (Category 1, Category 2, Category 3), 'Document types (4)' (Bill of Lading, Invoice, Utility Bill, Wage and Tax). The main area shows 'Category 3 sample documents (5)'. A search bar 'Search sample documents' is at the top. Below it is a table with columns 'Document name' and 'Upload'. The table contains five rows, each with a checkbox and a PDF file name: '✓ TR\_FW2\_1001\_0000\_P5.pdf', '✓ TR\_FW2\_2000\_0000\_P5.pdf', '✓ TR\_FW2\_3000\_0000\_P5.pdf', '✓ TR\_FW2\_3001\_0000\_P5.pdf', and '✓ TR\_FW2\_4000\_0000\_P5.pdf'. At the bottom 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, hover over the category name and **Click on the 3 dots** at the end of the category name.



The screenshot shows the 'Create document types' page in the IBM Cloud Pak | Automation interface. The left sidebar lists categories: 'Categories (3)' (Category 1, Category 2, Category 3), 'Document types (4)' (Bill of Lading, Invoice, Utility Bill, Wage and Tax). The main area shows 'Category 1 sample documents (2)'. A search bar 'Search sample documents' is at the top. Below it is a table with columns 'Document name' and 'Upload'. The table contains two rows, each with a checkbox and a PDF file name: '✓ UBillCache\_081\_1\_1\_1.pdf' and '✓ UBillCache\_082\_1\_1\_1.pdf'. The 'Category 1' label has a red box and three dots at its end. At the bottom right are 'Update categories', 'Back', and 'Finish' buttons.

8. Select ASSIGN TO DOCUMENT TYPE

The screenshot shows the 'Create document types' page in a web application. On the left, there's a sidebar with 'Categories (3)' containing 'Category 1', 'Category 2', and 'Category 3'. Below it is a list of 'Document types (4)'. In the main area, 'Category 1 sample documents (2)' is selected, and a list of PDF files is shown. At the top right, there are buttons for 'Update categories', 'Back', and 'Finish'. A red box highlights the 'Assign to document...' button in the sidebar.

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

The screenshot shows the 'Assign documents' dialog box. It has a title 'Assign documents' and a sub-instruction 'Assign documents of Category 1 to'. Two radio buttons are present: 'New document type' (unchecked) and 'Existing document type' (checked). A red arrow points to the 'Existing document type' radio button. Below this, a dropdown menu titled 'Utility Bill' is open, showing options like 'Bill of Lading', 'Invoice', 'Utility Bill' (which is selected and highlighted with a red arrow), and 'Wage and Tax'. At the bottom are 'Cancel' and 'Assign' buttons, with 'Assign' being blue.

10. Click Assign to close the dialog box.

11. Select the next Category 2 and Click on the 3 dots and Select Assign these documents 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

Assign documents

Assign documents of Category 2 to

New document type  Existing document type

Document type display name 18/50  
Mortgage Agreement

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

Document type symbolic name 17/50  
MortgageAgreement

Cancel Assign

\_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. Select Existing Document Type and Click Wage and Tax from the drop down and then Click on Assign.

Assign documents

Assign documents of Category 3 to

New document type  Existing document type

Document types

Invoice

Mortgage Agreement

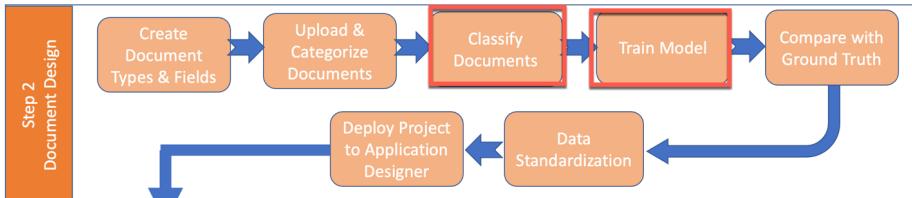
Utility Bill

Wage and Tax

Cancel 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. In the example below our project was called **<User01\_CEB>**
- \_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

This screenshot shows the 'Classification model' page in the IBM Cloud Pak | Automation interface. The accuracy is listed as 84.8%. The 'Test set (0)' section displays a message: 'There are no documents in the test set. Include at least 1 document in the test set to view training results.'

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

This screenshot shows the 'Classification model' page in the IBM Cloud Pak | Automation interface. The accuracy is listed as 84.8%. The 'Test set (2)' section now contains two documents: 'Mortgage Agreement1.pdf' and 'Mortgage Agreement2.pdf'. The 'Training set (3)' section contains three documents: 'Mortgage Agreement3.pdf', 'Mortgage Agreement4.pdf', and 'Mortgage Agreement5.pdf'.

- \_5. Select Wage and Tax on the Document types and move 2 documents over to the test set.



*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.*

## Automation Document Processing Lab

The screenshot shows the 'Classification model' page in the IBM Cloud Pak for Automation interface. At the top, it displays an accuracy of 84.8% and a note that the model was last trained 4 days ago. There are three buttons at the top: 'Back', 'Train' (which is highlighted in blue), and 'Review training results'. Below these, a message box 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, Invoice, Mortgage Agreement, Utility Bill, and Wage and Tax, with 'Wage and Tax' selected. The main area shows a split between a 'Training set (3)' containing 60% of total samples (with files TR\_FW2\_3000\_0000\_PS.pdf, TR\_FW2\_3001\_0000\_PS.pdf, and TR\_FW2\_4000\_0000\_PS.pdf) and a 'Test set (2)' containing 40% of total samples (with files TR\_FW2\_1002\_0000\_PS.pdf and TR\_FW2\_2000\_0000\_PS.pdf). A link to 'Auto generate 70/30 split' is available.

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

The screenshot shows the 'Classification model' page during the training process. The progress bar is at 30% complete, with about 21 minutes remaining. The 'Train' button is now greyed out. Other buttons like 'Back', 'Review training results', and 'Test trained model' are visible. A note says 'About 21 minutes remaining'.

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 yielded 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:

## Automation Document Processing Lab

The screenshot shows the IBM Cloud Pak for Automation interface. At the top, it displays 'Classification model' with an accuracy of 96.9%. Below this, there are buttons for 'Confirm inputs', 'Review samples', 'Review training results', and 'Test trained model (optional)'. A message box at the top right says 'Model trained successfully!' with the note 'Accuracy has been updated to reflect the latest changes.' In the center, a modal window titled 'Changes were made since you last trained your model. Retrain the model to get updated training results and accuracy.' is open. It contains a table titled 'Training results' with columns 'Document', 'Classified as', 'Classification result', and 'Confidence'. The table lists several documents, all classified as 'Bill of Lading' and marked as 'Correct', with confidence levels ranging from Medium to 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. 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

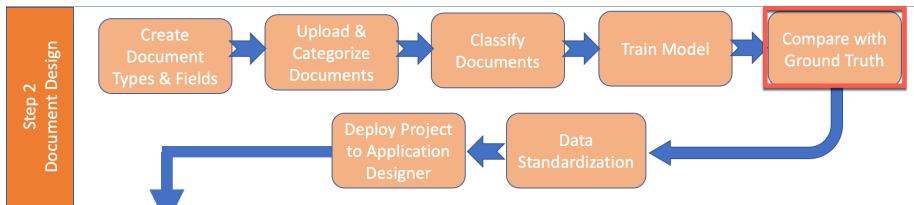
The screenshot shows the 'Document types and samples' section of the IBM Cloud Pak Administration interface. On the left, there's a sidebar with 'Document types' listed: Bill of Lading (23 samples), Invoice (26 samples), Mortgage Agreement (5 samples), Utility Bill (12 samples), and Wage and Tax (5 samples). The 'Wage and Tax' item is selected. The main area displays 'Wage and Tax sample documents (5)' with a search bar and a list of five PDF files. In the top right corner of this list, there is a blue 'Upload' button with a file icon, which is highlighted with a red box.

- \_4. From the zip files you downloaded and unzipped earlier upload all the files from the directory *Group 2 - Classification Results Increase Set*.
- \_5. Click on **Build** tab then let's retrain the Classification Module again.
- \_6. Click anywhere on **Classification model**.
- \_7. Click on **Wage and Tax**.
- \_8. Click **Next** button.
- \_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 / Carlos Baker Project'. It displays two lists of documents: 'Training set (14)' and 'Test set (6)'. Both lists show document names such as 'TR\_FW2\_1000\_0001F.pdf' and 'TR\_FW2\_2000\_0003F.pdf'. A yellow box at the top right says 'Please make sure you have at least 1 reviewed document to train the model.' Below the lists are 'Search' fields and 'Upload' buttons. At the bottom are 'Back' and 'Next' buttons.

\_3. 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.

\_4. 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.

\_5. Click on **Teach Samples**.

The screenshot shows the 'Teach model' step. The breadcrumb shows 'Business automation / User01-CB / Extraction model'. The 'Teach samples' button is highlighted with a red box. At the bottom are 'Upload' and 'Reanalyze' buttons.



*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.*

- \_ 6. 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 aliases we entered when we created the document class to locate data. Now, you need to correct and improve the model.

The screenshot shows the IBM Cloud Pak Administration interface with the following details:

- Document:** TR\_FW2\_1001\_0000\_PS.pdf
- Fields Extracted:**
  - Employee's social security number: 577-22-3048
  - Employer identification number: 14-023205
  - Employee name, address, and ZIP code: Michael Robert David Schubert, 56334 Full Street Avenue Unit 1234, Minneapolis, Minnesota 55411-1234
  - Address: 123456 A7B
  - Employee first name and initial: Michael
  - Employee last name: Robert
  - Employee middle name: David
  - Employee suffix: Schubert
  - Employee address and ZIP code: MN 123456789
  - State wages, tips, etc.: 123456789.99
  - State income tax: 123456789.99
  - Local wages, tips, etc.: 123456789.99
  - Local income tax: 123456789.99
  - Identify items: ABCDEFGH
- Extracted Fields Table:**

Field Name	Value Captured
Federal Income Tax Withheld	123456789.99
Social Security Wage	123456789.99
Medicare Wage	123456789.99
Social Security Tip	123456789.99
Medicare Tip	123456789.99
Social Security	123456789.99
Medicare	123456789.99
Alimony	123456789.99
Dependent care benefits	123456789.99
Nonqualified plan	123456789.99
Other	123456789.99
AAA 888 CCCC 12345678.90	123456789.99
AAA 888 CCCC 12345678.90	123456789.99
AA	123456789.99
- Actions:**
  - Show detected fields
  - Keyboard shortcuts on
  - Sort by: Date created
  - Save selection
  - Pending aliases | View all aliases (3)
  - None (0)
  - Mark this document as ready for training.



*Note: You may see different results than shown on the image above.*

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).

- \_1. ADP may have already preselected the first field but if not, **Click on the number below the heading "Federal Income tax withheld"** in the image.

The screenshot shows the IBM Cloud Pak Administration interface with a W-2 form document. The 'Federal Income Tax Withheld' field is highlighted with a red box. A floating panel titled 'Field Name Value Captured' shows the extracted value '1800.00'. Below it, a 'Recommended matches' section lists several options, with the first one ('Federal income tax 1800') having a green checkmark. A blue button labeled 'Save selection' is at the bottom right of the panel.

- \_2. 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**

Notice a green check mark signifies this field is complete.

The screenshot shows the IBM Cloud Pak Administration interface with a W-2 form document. The 'Federal Income Tax Withheld' field now has a green checkmark next to its value '123456789.99'. A blue button labeled 'Saved!' is at the bottom right of the extraction panel.

## Automation Document Processing Lab

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. Moving to Employee Name and Address field. Here it did pick up the address but missed the name. **Click on the Dismiss button**. Now select Draw button and using your mouse grab or lasso around “Employee’s first name and initial”. Again, **Click on Save match**

The screenshot shows a PDF document of a W-2 form on the left and a 'Recommended matches' dialog box on the right. The W-2 form contains various fields like Employee's social security number (577-22-3048), Employer identification number (14-023285), and employee information (Last name: Benjamin, First name: Charles, Address: 4326 Aldrich Rd, Minneapolis, MN 55412). The 'Recommended matches' dialog lists several suggestions for the employee's first name and initial, with one item selected: 'e Employee's first name and initial' (4326 Aldrich Rd, Minneapolis, MN 55412). The dialog also shows other suggestions for the employee's address and ZIP code.

The field label has been populated but we still need the field value.

- \_4. You will see that there are a series of blue underlines below all the characters found. We can see that ADP wasn’t able to pick up all the characters accurately. We are interested in getting the “Employee’s First Name” data and the field value. **Click on the Draw button** under Field value. Using your mouse select the Field Value then **Click on the Draw button** under Field Value and select **Name and**

address (green box) along with the address, then **Click Save selection**

The screenshot shows the IBM Cloud Pak Administration interface. On the left is the W-2 Wage and Tax Statement form for 2020. On the right is a field extraction tool window. The tool highlights several fields with red boxes: 'Employee's first name and initial' (containing 'Benjamin P. Charles') and 'Employee's last name' (containing 'Charles'). It also highlights the address '4326 Aldrich Rd Minneapolis, MN 55412'. A 'Save selection' button is highlighted with a red box at the bottom right of the tool window.

- \_5. For the Employee Social Security field if it looks good, **Click on Save selection**.
- \_6. Continue to process for the remaining fields, using either method as described above, clicking on the Save selection if correct or select *Dismiss* and use blue lines if Key Value Pair (KVP) is correct or drawing a box around needed label or value.
- \_7. Once complete **check the box** next to “Mark this document as ready for training” at the bottom

This screenshot is similar to the previous one but includes a red arrow pointing to the 'Mark this document as ready for training' checkbox at the bottom of the 'Save selection' dialog. The checkbox is checked, indicating the document is marked as ready for training.



- \_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 one where there is nothing but blue dotted lines. These blue lines are showing the characters that ADP did pick up. For example:

The screenshot shows the IBM Cloud Pak Administration interface with two W-2 Wage and Tax Statement forms from 2019. Both forms show identical data:

- Employee's name, address, and ZIP code:** Long Lengthy Name The Corporation, 56334 Full Sized Avenue Unit 1234, Minneapolis, Minnesota 55411-1234
- Employer's name, address, and ZIP code:** Michael Robert David Smithson III, 56334 Full Sized Avenue Unit 1234, Minneapolis, Minnesota 55411-1234
- State:** MN
- Social Security No.:** 123456789
- State Income Tax No.:** 123456789.99
- Federal Income Tax Withheld:** 123456789.99
- Dependent care benefits:** 123456789.99
- Nonqualified plans:** 123456789.99
- State Income Tax Withheld:** 123456789.99
- Employer Identification Number (EIN):** 14-023285
- Employee's Social Security No.:** 123456789.99
- Employee's Date of Birth:** 57-22-2048
- Local Income Tax:** AAA BBB CCC
- Local Income Tax:** 123456789.90
- Local Income Tax:** 123456789.99
- Local Income Tax:** ABCDEFGH
- Copy B To Be Filed With Employee's FEDERAL Tax Return:** This information is being furnished to the Internal Revenue Service. Only No. 1040 is filed.
- Dept. of the Treasury - IRS:** Verifying IRS website at www.irs.gov

To the right of the forms is a "Field capture" tool window:

- Sort by:** Date created
- Field Name:** Federal Income Tax Withheld
- Value Captured:** 123456789.99
- Field Type:** Text
- Field Label:** Draw (highlighted)
- Field Value:** Draw (highlighted)
- Pending changes:** None
- Save selection:** Save selection button
- Actions:** Employee Name and A..., Employee Social Secur..., Employer Identification..., Employers Name and A... buttons
- Mark this document as ready for training:** Checkmark box
- Buttons:** Previous sample, Next sample

By simply clicking on the Field value it will populate both the label and field value with a pop-up window asking if you want to save the match.

## Automation Document Processing Lab

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

The screenshot shows the JHM Cloud Pak Administration interface with two sample W-2 forms and their corresponding training annotations.

**Form W-2 Wage and Tax Statement** (Copy 2 - For State, City or Local Tax Department) Year 2020 OMB No. 1545-0008

1. Employee Social Security Number: 208-47-1017	2. Employee ID number: 98-749972	3. Control number: AAKH45
4. Employee Name & Address: David Gomez Top Heights Markets LLC 123 Main Street New York, NY 10007	5. Employer Name & Address: Top Heights Markets LLC 123 Main Street New York, NY 10007	6. Control number: AAKH45
7. Federal income tax withheld: \$10000.00	8. State income tax withheld: \$0400.00	9. Social security tips: \$8239.80
10. Medicare tips: \$275.00	11. Nonqualified Plans: \$0	12a. Code (see instructions): \$423.00
12b. Dependent care benefits: \$0	13. Other: D	14. Other: \$2000.00
14. (Mark if) Statutory employee: X	15. Other: D	16. Code: \$32.00
16. State ID-number: NY	17. State wages, tips, etc.: 795007	18. State code: 425.00
17. State income tax: \$14700.00	18. Local wages, tips, etc.: 210000.00	19. Local code: Red Beach

**Form W-2 Wage and Tax Statement** (Copy 2-To Be Filed with Employee's State, City or Local Income Tax Return) Year 2020 OMB No. 1545-0008

1. Employee Social Security Number: 208-47-1017	2. Employee ID number: 98-749972	3. Control number: AAKH45
4. Employee Name & Address: David Gomez Top Heights Markets LLC 123 Main Street New York, NY 10007	5. Employer Name & Address: Top Heights Markets LLC 123 Main Street New York, NY 10007	6. Control number: AAKH45
7. Federal income tax withheld: \$21000.00	8. State income tax withheld: \$0400.00	9. Social security tips: \$8239.80
10. Medicare tips: \$275.00	11. Nonqualified Plans: \$0	12a. Code (see instructions): \$423.00
12b. Dependent care benefits: \$0	13. Other: D	14. Other: \$2000.00
14. (Mark if) Statutory employee: X	15. Other: D	16. Code: \$32.00
16. State ID-number: NY	17. State wages, tips, etc.: 795007	18. State code: 425.00
17. State income tax: \$14700.00	18. Local wages, tips, etc.: 210000.00	19. Local code: Red Beach

**Form W-2 Wage and Tax Statement** (Copy 3 - To Be Filed with Employee's FEDERAL Tax Return) Year 2020 OMB No. 1545-0008

1. Employee Social Security Number: 208-47-1017	2. Employee ID number: 98-749972	3. Control number: AAKH45
4. Employee Name & Address: David Gomez Top Heights Markets LLC 123 Main Street New York, NY 10007	5. Employer Name & Address: Top Heights Markets LLC 123 Main Street New York, NY 10007	6. Control number: AAKH45
7. Federal income tax withheld: \$21000.00	8. State income tax withheld: \$0400.00	9. Social security tips: \$8239.80
10. Medicare tips: \$275.00	11. Nonqualified Plans: \$0	12a. Code (see instructions): \$423.00
12b. Dependent care benefits: \$0	13. Other: D	14. Other: \$2000.00

**Form C - For Employee's Records** Year 2020 OMB No. 1545-0008

1. Employee Social Security Number: 208-47-1017	2. Employee ID number: 98-749972	3. Control number: AAKH45
4. Employee Name & Address: David Gomez Top Heights Markets LLC 123 Main Street New York, NY 10007	5. Employer Name & Address: Top Heights Markets LLC 123 Main Street New York, NY 10007	6. Control number: AAKH45
7. Federal income tax withheld: \$21000.00	8. State income tax withheld: \$0400.00	9. Social security tips: \$8239.80
10. Medicare tips: \$275.00	11. Nonqualified Plans: \$0	12a. Code (see instructions): \$423.00
12b. Dependent care benefits: \$0	13. Other: D	14. Other: \$2000.00

The right side of the interface shows a list of extracted fields with their corresponding values and annotations:

- Federal income tax withheld: 210000.00
- Employee Name and A... Required: David Gomez
- Employee Social Securi... Required: 98-749972
- Employer Identification...: 328-47-1017
- Employers Name and A...: Top Heights Markets LLC 563 3rd...
- Social Security Wages: 132099.00
- Wages Tips Other Com...: 210000.00

Annotations include:
 

- Annotations for "Federal income tax withheld" and "Employee Name and A..." are marked with green dots.
- Annotations for "Employee Social Securi...", "Employer Identification...", "Employers Name and A...", "Social Security Wages", and "Wages Tips Other Com..." are marked with blue dots.
- A checkbox "Mark this document as ready for training" is checked.

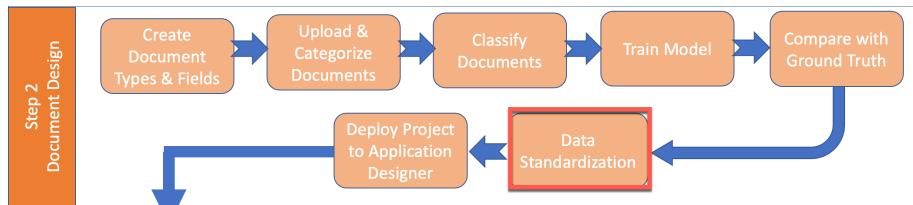
## 8.2 Train extraction model

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



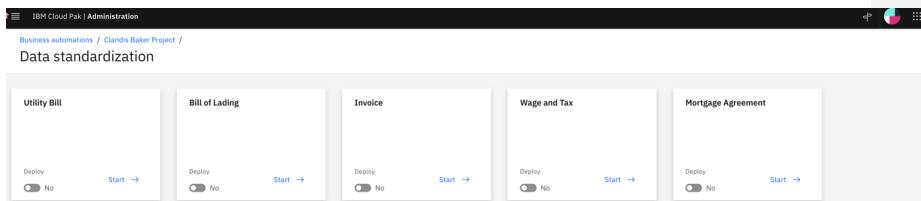
Next, we may need to standardize the data that will be presented in the user interface and how it will be stored in the FileNet repository for example. Data standardization is the process of defining attributes for a data field in a standardized way. This is done using data definitions. These definitions can be used across projects, and across different applications within the CloudPak 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

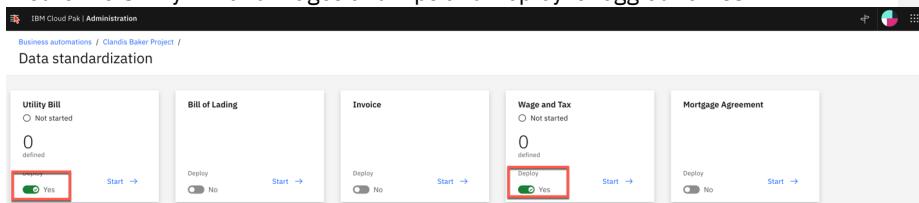
Category	Status	Count	Details
Document types and samples	Ready	4 types	samples on average
Classification model	Ready	3 types trained	100% accuracy
Extraction model	Ready	3 types trained	97% accuracy
Data standardization	Not ready		

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



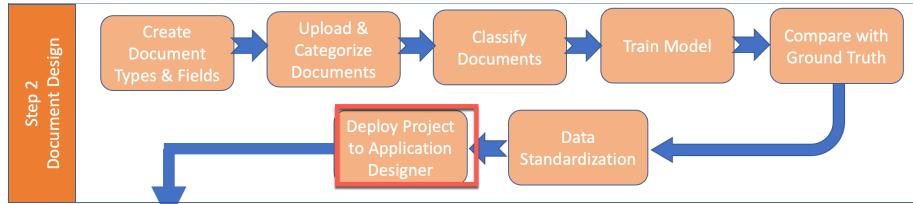
\_3. Click on Start on either selected deployment.

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.

[Business automations / Clandis Baker Project /](#)

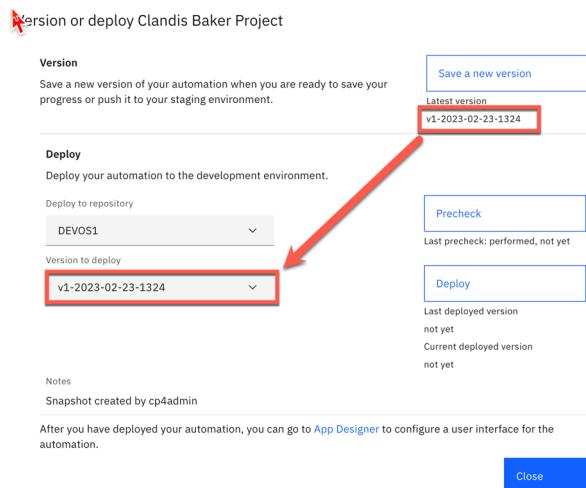
## 10 Version and deploy your project



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.**
- \_3. Once the version is saved, you should see the version in the Version to deploy drop down list

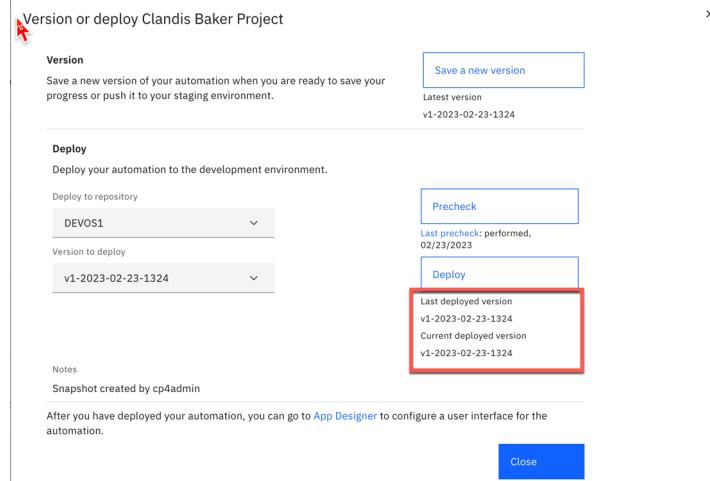


... also, in the top corner has the “Latest Version.”

## Automation Document Processing Lab

- \_4. Click on the **Deploy button**. This will also take several minutes and potentially time out if others are also trying to deploy.

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

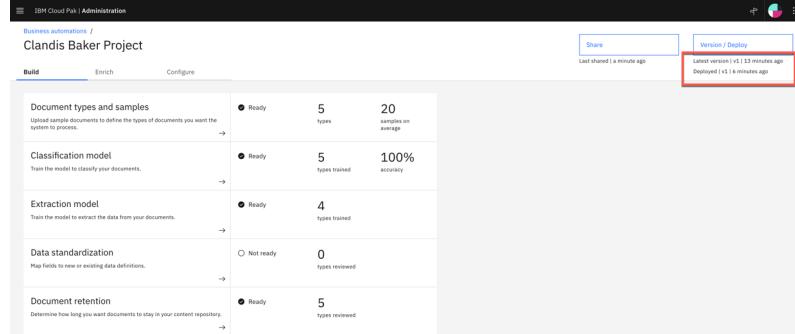


Note that you do not have to remain in the deploy screen while it is versioning or deploying. You can always click the button and then go back into any other screen if you like. It will run in the background. If you do this, just keep an eye on the top right of your screen for deployment status.

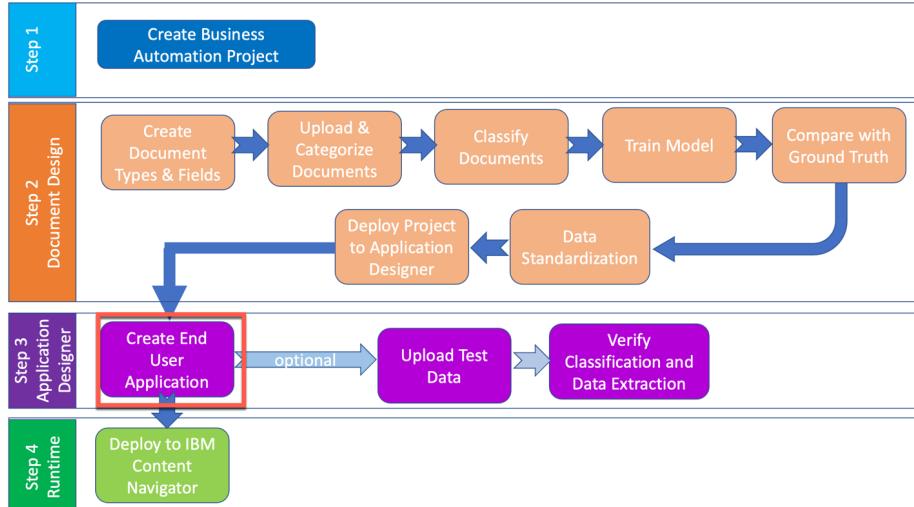
- \_5. Click **Close** button.

Once deployed, proceed to the next steps.

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



## 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.

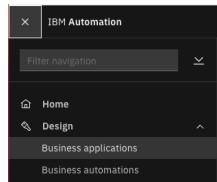
### 11.1 Create your Runtime Application.

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



and selecting **Business Applications**

## Automation Document Processing Lab



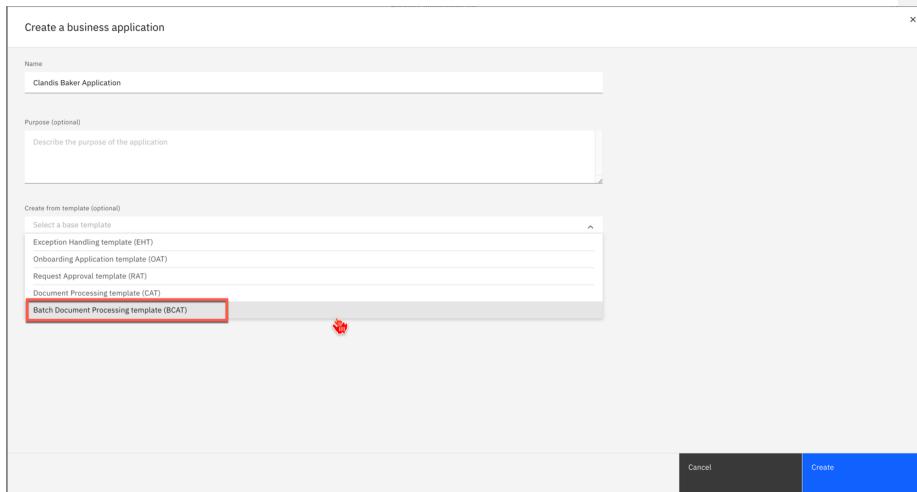
\_2. From the **Create** drop down list, select Application

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)', the 'Select a base template' dropdown is open, showing options like 'Exception Handling template (EHT)', 'Onboarding Application template (OAT)', etc. At the bottom 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**

Create an application

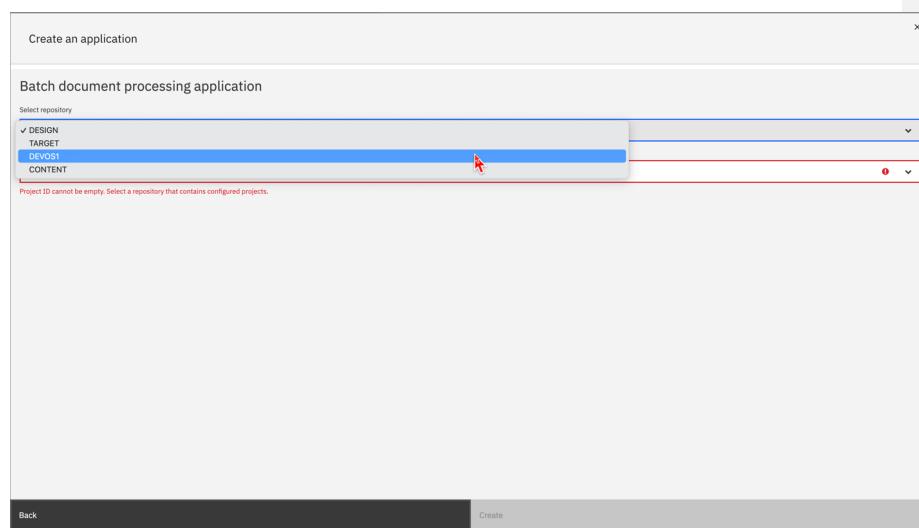
Batch document processing application

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

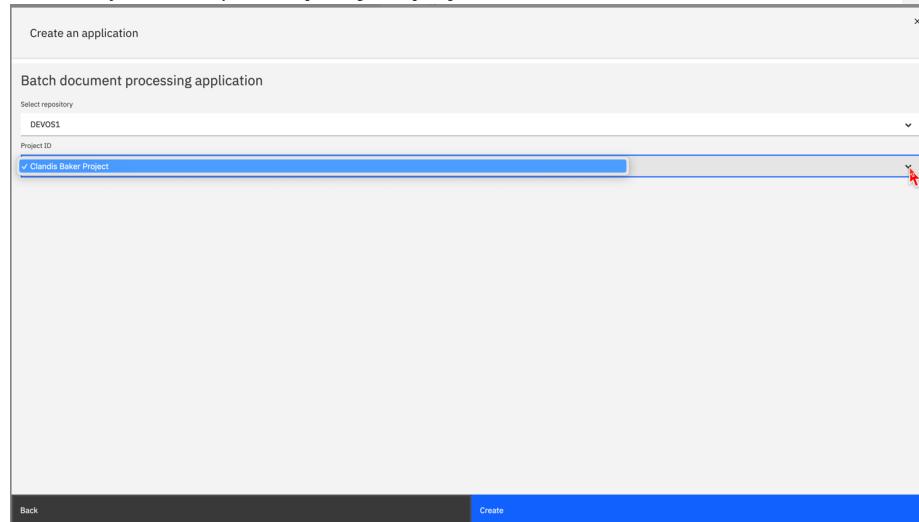
Select repository

DEVOVS1

Project ID

✓ Clandis Baker Project

Back Create

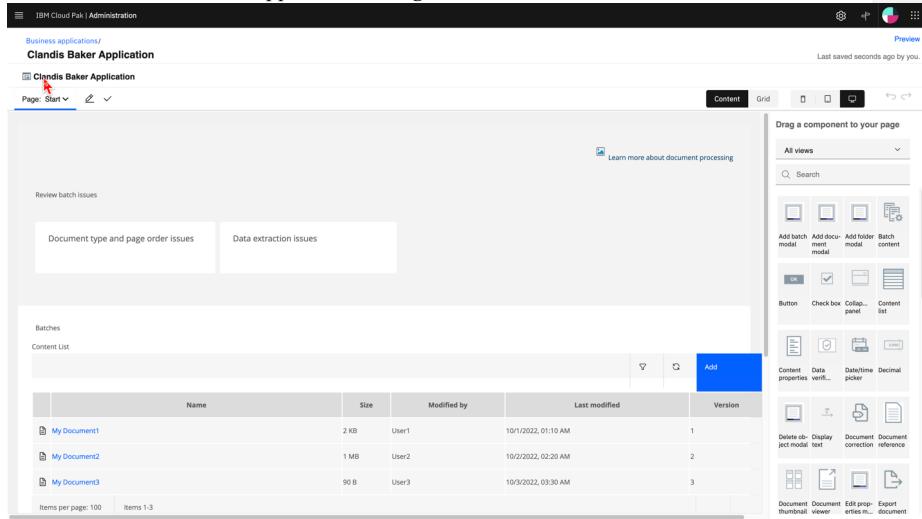


*Note it may take a minute or two before this update and you can see your project*

## 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. The top navigation bar includes 'IBM Cloud Pak | Administration', 'Business applications / Cländis Baker Application', and 'Preview'. The main area is titled 'Cländis Baker Application' with a sub-section 'Page: Start'. On the left, there's a sidebar with 'Review batch issues' and two tabs: 'Document type and page order issues' and 'Data extraction issues'. The main content area displays a 'Batches' section with a 'Content List' table. The table has columns: Name, Size, Modified by, Last modified, and Version. It contains three rows: 'My Document1' (2 KB, User1, 10/1/2022, 01:10 AM, Version 1), 'My Document2' (1 MB, User2, 10/2/2022, 02:20 AM, Version 2), and 'My Document3' (90 B, User3, 10/3/2022, 03:30 AM, Version 3). An 'Add' button is located at the top right of the table. To the right of the table is a sidebar titled 'Drag a component to your page' with sections for 'All views', 'Search', and various component icons like 'Add batch modal', 'Button', 'Content properties', etc.



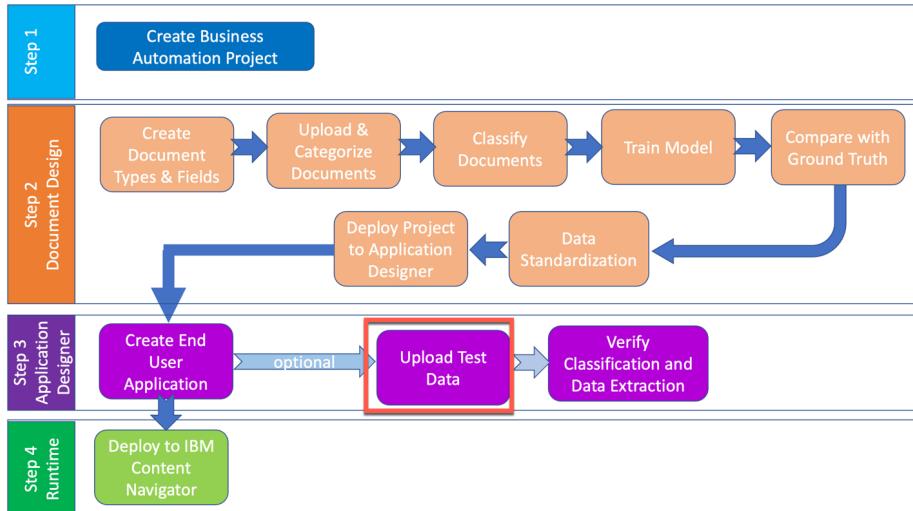
Batch Document Processing template (BCAT) has all the necessary pages and configuration to start using the application. Using this designer user interface, you have the option to further customize the application, such as its page design or actions, to fit your requirements.

### \_9. Click **Preview** at the top right corner.



*Note: It may take several seconds to build and display the current configuration of the interface.*

## **11.2 Upload documents for processing**



\_1. You should be in the default application user interface for ADP.

IBM Cloud Pak | Administration > Claudi Baker Application > +

Not Secure | https://cpd-cp4ba-starter.apps.ocinstall.gym.lan/ie-pbx/Claudi%20Baker%20Application/CBA/?loc=en

Bookmarks this! w3 Doc Imaging and... CvC - Home - Dat... Customer Log In... Sign In - Skytap WW SWAT Seismic - Login ECM Enablement... Advanced Case M... ADP IBM GitHub Log In - Jazz Team... Other Bookmarks

Learn more about document processing

### Review batch issues

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

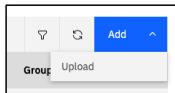
### Batches

Name	Files	Priority	Status	Added on	Added by	Group	Location	Add
								

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.

Upload new batch

\* Display Name  
Batch 1

Description

Priority  
High

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

Navigate to the samples folder previously downloaded and use the *Group 3 - Runtime 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.

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.

File Name	Document Type
B_PO_5.pdf	Auto Classify
DE_FW2_1000_0001F.pdf	Auto Classify
DE_FW2_4000_0011F.pdf	Auto Classify
DE_FW2_4001_0001S.pdf	Auto Classify
DE_FW2_4001_0010F.pdf	Auto Classify

1 items selected

Classify | Auto Classify | Deselect

Cancel Add

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 boxes: 'Document type and page order issues' (0 batches) and 'Data extraction issues' (0 batches). Below is the 'Batches' section with a table. A progress bar at the top of the table indicates '3 of 5 files processed'. A red arrow points to the 'Add' button at the top right of the table header.

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		

Review batch issues

Document type and page order issues 0 batches

Data extraction issues 0 batches

Learn more about document processing

Batches

Items per page: 100 1-1 of 1 items

Add

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 same 'Review batch issues' and 'Batches' sections as the previous step. A red arrow points to the three-dot menu icon at the end of the table header row.

Name	Files	Priority	Status	Added on	Added by	Group	Location
Batch01	5	High	<div style="width: 100%;">Documents uploaded</div>	02/23/2023, 10:49 AM	cp4admin		

Review batch issues

Document type and page order issues 0 batches

Data extraction issues 0 batches

Learn more about document processing

Batches

Items per page: 100 1-1 of 1 items

### \_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.

Name	Files	Priority	Status	Added on	Added by
Batch 1	6	High	Document issues	01/13/2021, 08:44 am	CEAdmin

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.

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

Batch1

The screenshot shows a list of documents in a 'Batch1' folder. One document, 'B\_PO\_5.pdf', is highlighted with a red border. To its right is a detailed view of a 'PURCHASE ORDER' from 'RUBE'S Meat Co.'.

**PURCHASE ORDER**

**RUBE'S Meat Co.**

**VENDOR:** Chicken Run Ranch  
24 Quarry Street  
Helen Village NE23 6DD  
078-2054-8486

**SHIP TO:** Rube's Meat Co.  
14 Market Road  
Burton, Leonard HG3 2SU  
UK  
078-7875-2017

SHIPPING METHOD		SHIPPING TERMS		DELIVERY DATE	
AIR		C.I.F.		29 March 2020	
QTY	ITEM #	DESCRIPTION	JOB	UNIT PRICE	LINh TOTAL
200 PCS	01	Whole Chicken		£1.50	£340.00
100 Packs	02	One Day Old Chick		£1.00	£107.50
					TOTAL £602.50

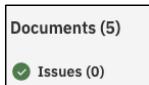
Commented [CB3]: May need new screen

- \_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.

Batch1

The screenshot shows the same batch of documents. The trash can icon next to the purchase order document ('B\_PO\_5.pdf') is highlighted with a red box, indicating it is selected for deletion.

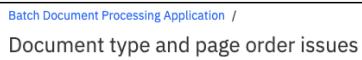
- \_4. Review all documents to ensure everything is correct. If the system no longer detects any issues, you should see a green checkmark near the top of the document list.



\_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 **Batch Document Processing Application** link at the top to return to the previous preview menu.



#### 11.4 Correct extraction issues

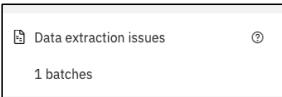
The following instructions are based on a pre-trained sample application. Not what you will see in your untrained application.



*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.



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

## Automation Document Processing Lab



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

A screenshot of a web application interface titled 'Batch Document Processing Application / Batches with data extraction issues / Batch1'. The page shows a table of documents with their status, modified date, and by whom. One document, 'DE\_FW2\_4001\_0010F.pdf', is highlighted with a red border and has a yellow triangle icon next to it, indicating it has data issues. The table has columns for Name, Issues, Status, Modified on, and Modified by. A 'Submit' button is visible at the top right. At the bottom left, it says 'Items per page: 100' and '1-4 of 4 items'.

Notice one of the documents has Data issues.

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

A screenshot of a tax form viewer. On the left, a preview of the tax form 'Form W2 Wage and Tax Statement' is shown with a blue border around the header. On the right, a detailed view of the form fields is displayed. A red arrow points to a yellow warning icon in the top right corner of the viewer window. Below the warning, a panel titled 'Issue types' lists various extracted data fields such as Employee Name and Address, Organization, Name, Email, Phone, and Postal mail address, each with a '(none)' status indicator.

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. At the top, there's a toolbar with icons for zoom, rotate, and other document operations. Below the toolbar is a preview area showing a tax form (Form W2) and its extracted data. The extracted data is displayed in a sidebar on the right, organized into sections like Employee Name and Address, Name, Email, Phone, and Postal mail address. The extracted data includes fields such as Employee ID number (87-3849583), Control number (EED64 FGH), Social security wages (132099.00), and Social security tax withheld (83239.80). The tax form itself contains various fields like Employee social security number, Employer name & address, and various wage and tax amounts.

- Rotate image.
- Visual effect adjustment
- Invert

### Image viewer features at bottom:

This screenshot shows the same document processing interface as the top one, but with a different view of the image viewer. The preview area at the bottom now has a red box highlighting the page and thumbnail navigation buttons. The rest of the interface, including the extracted data sidebar and the tax form content, remains the same.

- Page and thumbnail's view
- Fit to window.
- Zoom and Magnify

## Automation Document Processing Lab

### Field features

The screenshot shows a document titled "Form W-2 Wage and Tax Statement" from the Internal Revenue Service. The document includes fields for Employee social security number (338-91-3066), Employer ID number (87-3849583), Control number (EE2640 FGH), and various wage and tax amounts. To the right, a "Extracted data" panel is open, showing a dropdown menu with "Fields with issues" selected. The "Fields with issues" section highlights several fields with validation icons (yellow triangles). The panel also displays sections for Employee Name and Address, Organization, Name, Email, Phone, and Postal mail address.

- Show all fields.
- Show fields with issues.

Also note that fields that do have issues have a notification icon next to them. For example, if the Employee Social Security Number field is a mandatory field and expects a numeric value. But in this this field also has hyphens in it wouldn't pass validation.

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

A screenshot of a dropdown menu titled "Extracted data". It contains two options: "All Fields" (selected) and "Fields with issues". A red arrow points to the "Fields with issues" option.

\_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.



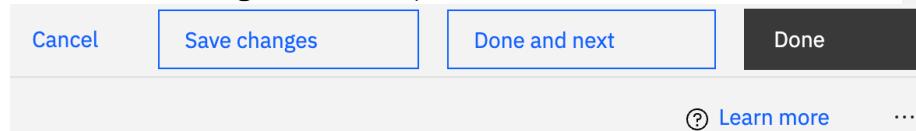
*The Employee Name and Address is a mandatory field. It is also a composite field. Remember when we defined this field we picked Address Information, because we wanted not only the Address block but also the name.*

*This is why you see multiple fields under the Employee Name*

\_6. Click on Name and with your mouse select the person's name.

The screenshot shows a document processing interface with a W2 form on the left and its extracted data on the right. The extracted data panel has a tree view under 'Employee Name and Address'. The 'Name' node is expanded, showing 'Name: Francis Hallbut'. A red arrow points to this name field.

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



\_8. For the remaining fields there are no extraction issue only low confidence characters. Click on Dismiss for each field with a yellow validation warning.

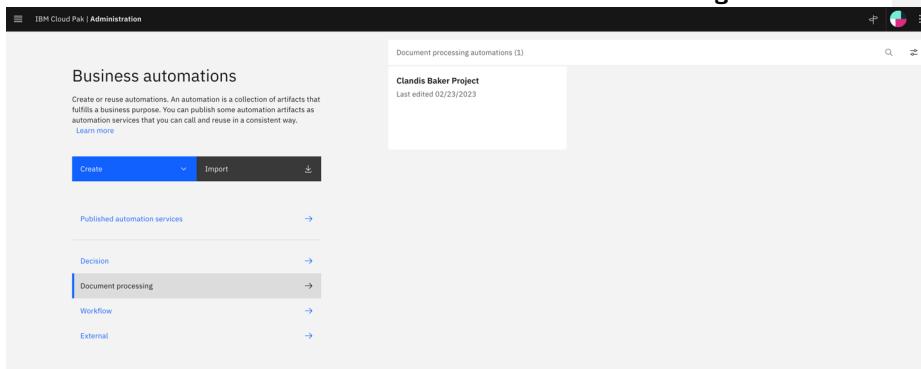
\_9. Click on Done and next..

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

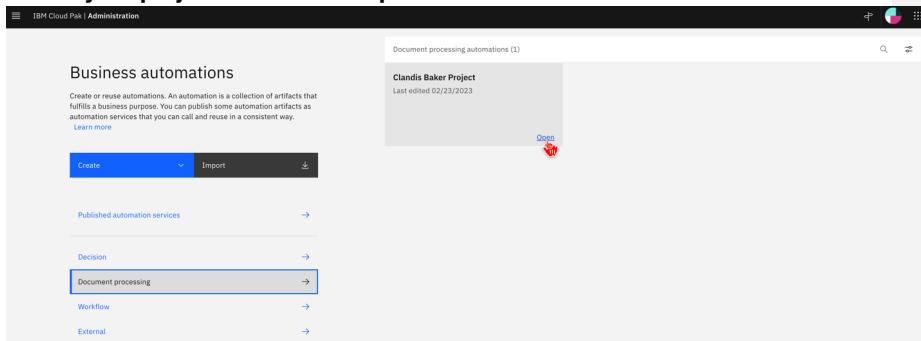
## 12 Optional Export/Import Project.

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**.

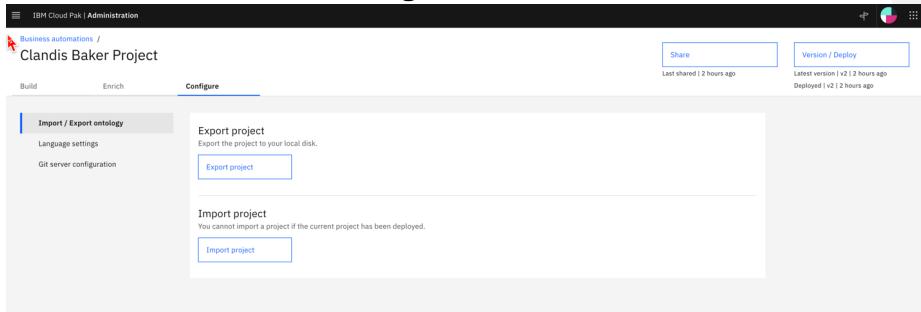


- \_2. Select <your project name> Click open



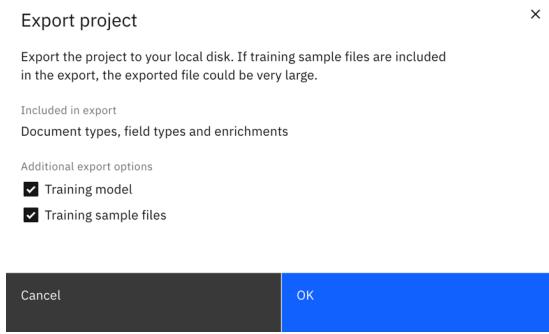
## Automation Document Processing Lab

\_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.

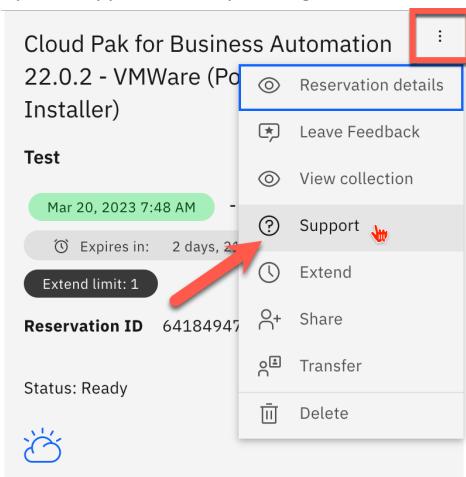
An operator fails to install and continuously shows Pending status.

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.

### Can't find user/password in Daffy

If your deployment has FAIL when looking into getting username and password then your environment is not working.

## Automation Document Processing Lab

```
#####
# Daffy Options #
#####
Please use this tool and select what option you would like to retrieve more info on.
With this menu you can get your OpenShift Console URL, id/passwords and status.
You can also get your CP4BA Console URL, id/passwords and status info.

1) OpenShift
2) Services
3) MainMenu
#? 2
CP4BA Services Menu:
1) Console
2) Status
3) Back
#? 1
#####
Running daffy service process v2023-01-11
Log File - /data/daffy/log/ocpinstall/cp4ba/service.sh-2023-03-05-10-47.log
#####
Start time : Sun Mar  5 10:47:01 EST 2023

Checking OS before continuing on
#####
Linux is being used (Red Hat Enterprise Linux 8.7 (Ootpa))

Login via oc(ocpadmin)
#####
oc login https://api.ocpinstall.gym.lan:6443 -u ocpadmin -p ***** --insecure-skip-tls-verify
admin user - ocpadmin

Validate OCP Access
#####
✓ PASSED Access to cluster via oc command

Validate CP4BA version info
#####
✓ PASSED Valid version CPBA_VERSION=22.0.2

Console Automation Document Processing
#####

Daffy Version          : v2023-01-11
Bastion OS             : rhel - 8.7
Platform Install Type : vsphere-ipi
OpenShift Cluster Name: ocpinstall
OpenShift Version      : 4.10.36
CP4BA Version          : 22.0.2
Project/Namespace     : cp4ba-starter
Zen Version            : 4.8.0
Message 1              : Running reconciliation
Message 2              : Prerequisites execution done.
Message 3              : FAIL - prerequisites Deployment failed ←
Message 4              :
Deployment Service    : Starter docprocessing
Config Map Dump        : /data/daffy/log/ocpinstall/cp4ba/icp4adeploy-cp4ba-access-info.yaml
```

### **\*\*\*Environment verification\*\*\***

Once you have reserved a cluster in IBM TechZone, it is first **\*\*Scheduled\*\*** for provisioning. After a while it moves into status **\*\*Provisioning\*\***, and after some time finally becomes **\*\*Ready\*\***.

At that time, you'll also get an email that your cluster is Ready, but this only means that the Red Hat OpenShift part is now available. Once the cluster is Ready, the deployment of the CP4BA Starter pattern will automatically be performed. Therefore, you must wait until not only the OCP cluster has been provisioned but also until CP4BA Starter pattern has been completely deployed.  
**\*\*\*Combined this may take several hours (~5-6 hours).\*\*\***

At the moment, there is a known Red Hat OpenShift bug that can intermittently block the successful deployment of CP4BA Starter pattern. To identify that your TechZone provisioned environment has hit this issue, **please check about one hour after the cluster has become ready** if your cluster is affected by this bug.

For this, please perform the following steps:

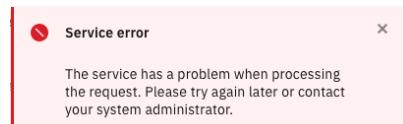
- Open the *OpenShift web console* in a browser.
- In the left-hand side navigator go to *Operators -> Installed Operators*.
- Make sure the *project scope* is set to *All Projects*.
- Verify that *all Operators* show in the column with *Status* the value *Succeeded*.
- If there are one or multiple Operators *NOT with Status 'Succeeded'* (for example in Status 'Failed', 'Unknown', or 'Cannot update'), your environment is affected by the mentioned bug and applying a manual workaround is required. For this, please reach out for [Support](#support).
- Once all Operators show in column *Status 'Succeeded'*, you can proceed with the next prerequisite.

To verify that your CP4BA cluster is completely deployed:

- Open the \*\*OpenShift web console\*\* in a browser.
- Click on \*\*Workloads -> ConfigMaps\*\* on the left-hand side navigator.
- Type \*\*\*access-info\*\*\* in the field next to 'Name'.

If the ConfigMap \*\*\*icp4adeploy-cp4ba-access-info\*\*\* is shown, your CP4BA cluster is deployed.

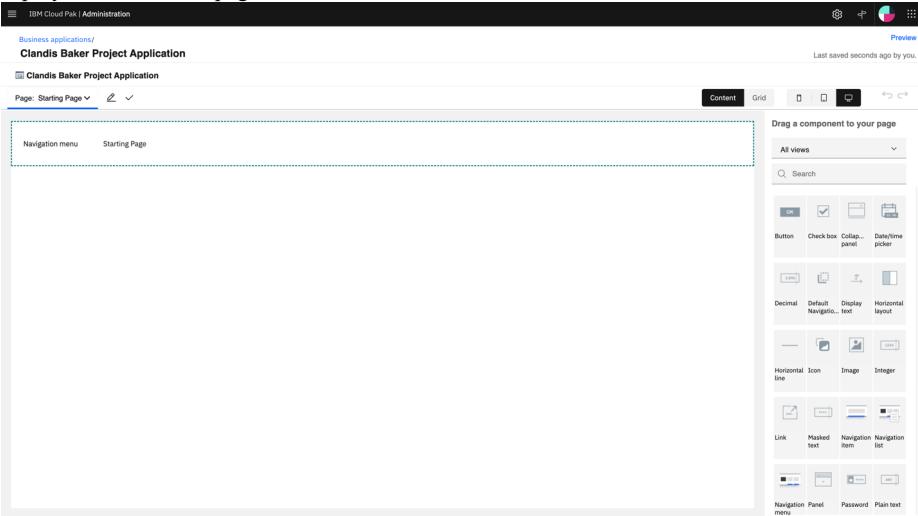
## Service Error



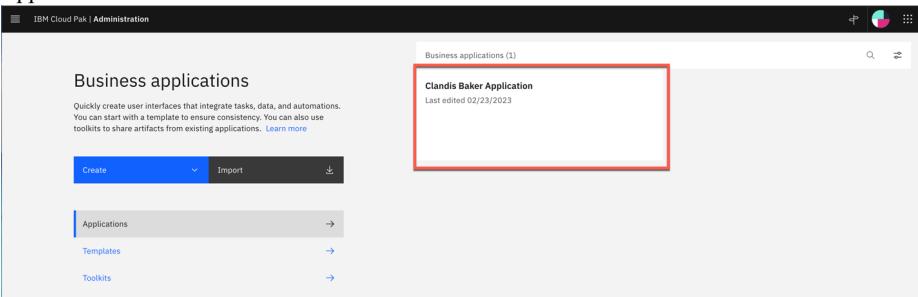
There was trouble communicating most of the time you can simply ignore and continue on.

## Application Blank

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

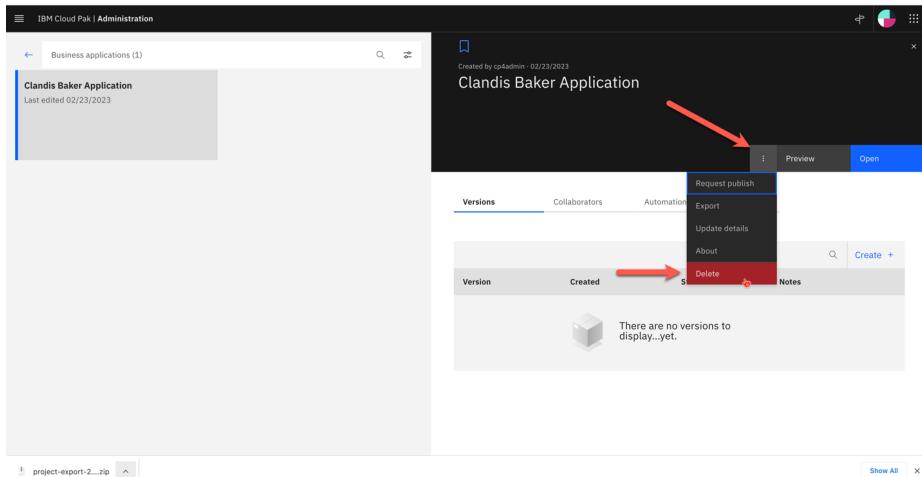


If this happens delete the application and try again. To delete the application, Click on the Application tile



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.

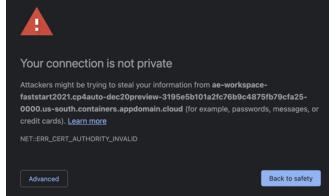
### Opening an Incognito Window

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:

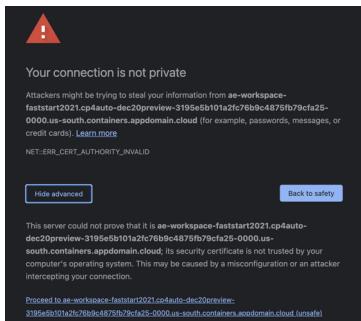
Open the Generate Security Tokens Box note and click all 3 of the links listed. This will reset the self-signed security certificates.

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

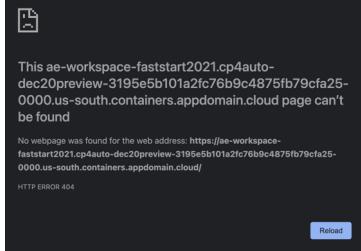


Click Advanced, and the browser window will look something like this:

## Automation Document Processing Lab



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

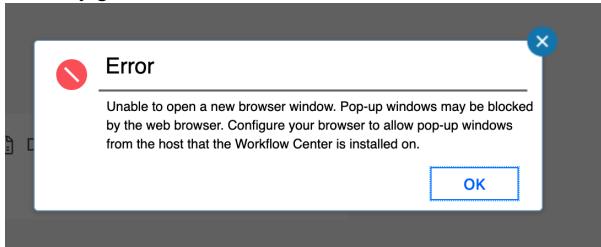


Ignore the error and proceed to the next link.

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

<https://github.com/IBM/baw-adp-integration-sample>

## Appendix C - Badge Information.

Badge quiz page - <https://learn.ibm.com/course/view.php?id=12413>

Credly page - <https://www.credly.com/org/ibm/badge/ibm-automation-document-processing-tech-jam>