

# IBM Cloud Pak for Business Automation Demos and Labs 2025

Documents Processing with DataCap+Watsonx.ai  
in Cloud Pak for Business Automation

V 2.0.1 (for CP4BA 25.0.0)

Sundeep Anne  
[Sundeep.Anne@ibm.com](mailto:Sundeep.Anne@ibm.com)

Pooja Luthra  
[pooja.luthra@ibm.com](mailto:pooja.luthra@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 of 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.

## Contents

<b>1.</b>	<b>Introduction.....</b>	<b>5</b>
1.1	IBM Datacap .....	5
1.2	Lab Overview .....	5
1.3	Overview of Exercises.....	6
1.4	Audience .....	7
1.5	Getting HELP during the lab .....	7
<b>2.</b>	<b>Lab Setup.....</b>	<b>8</b>
2.1	Provision Techzone Environment.....	8
2.1.1	Reserve Environment .....	8
2.2	Getting Started – Accessing the Environment .....	16
2.3	Introducing the components .....	18
<b>3.</b>	<b>Exercises – Configuration Lab Instructions.....</b>	<b>20</b>
3.1	Exercise 1: Create Datacap application by using Datacap FastDoc Admin .....	20
3.1.1	Datacap FastDoc – Introduction .....	20
3.1.2	Datacap FastDoc – Step by Step Instructions .....	20
3.1.3	Datacap FastDoc – Verification Instructions .....	27
3.1.4	Datacap FastDoc – Summary .....	29
3.2	Exercise 2: Enhance and complete the Datacap application by using Datacap Studio.....	29
3.2.1	Datacap Studio .....	29
3.2.2	Datacap Studio – Step by Step Instructions .....	29
3.3	Exercise 3: Datacap Navigator (optional).....	88
3.3.1	Datacap Navigator System Overview .....	88
3.3.2	Datacap Navigator Introduction – Step by Step Instructions .....	88
3.3.3	Overview of new Datacap Desktop .....	97
3.3.4	RuleRunner Setup .....	101
3.3.5	Scanning Documents .....	104
3.3.6	Verify a Batch .....	107
3.3.7	Viewing Results in Repository.....	109
4.	Troubleshooting .....	111
5.	Explore Document Redaction with Watsonx.ai .....	114
6.	Reference Links: .....	119

# 1. Introduction

## 1.1 IBM Datacap

IBM Datacap is a complete solution for document and data capture. Datacap scans, classifies, recognizes, validates, verifies, and exports data and document images quickly, accurately and cost effectively.

By combining the common recognition engines for Optical Character Recognition (OCR), Intelligent Character Recognition (ICR), Optical Mark Reading (OMR) and barcodes with libraries of hundreds of script-based and code-based (.NET) actions, Datacap accurately captures data from any type of structured, highly variable, or unstructured documents.

Datacap can capture machine print, handprint, bar codes, and check box data. By using the Datacap rules engine, data capture can be tailored to fit the most demanding business requirements and can be changed quickly when business needs change.

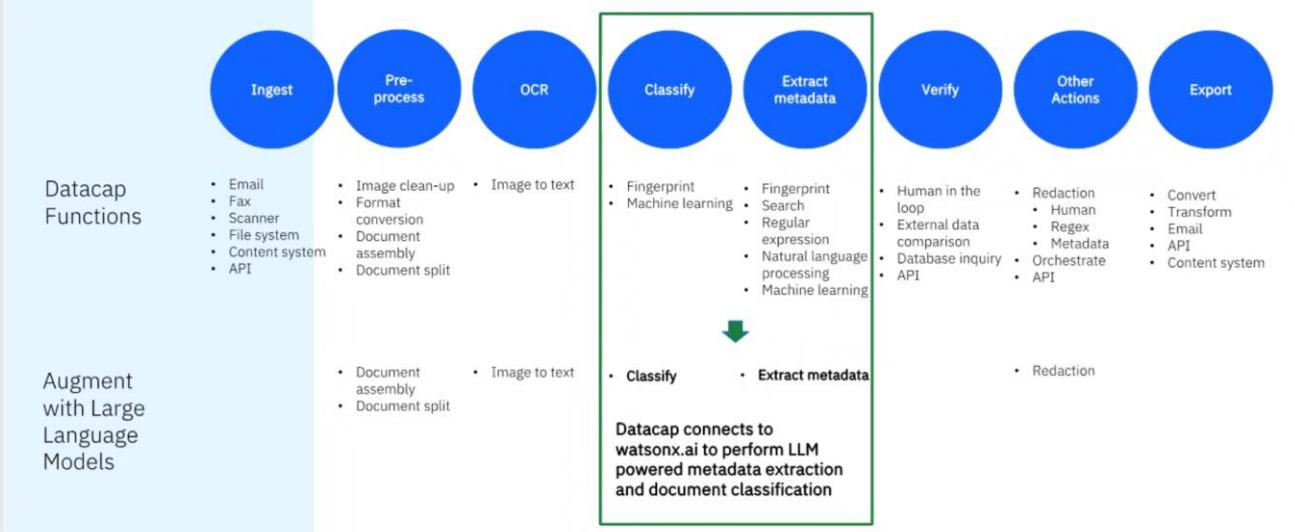
This scenario describes how to create and configure an application in DataCap and to integrate with Watsonx.ai using two authoring tools (**FastDoc Admin and Datacap Studio**), connect it to an **ECM system** as well as use the created application with **Datacap Desktop** (Watsonx client) and **Datacap Navigator** (web based through IBM Content Navigator).

## 1.2 Lab Overview

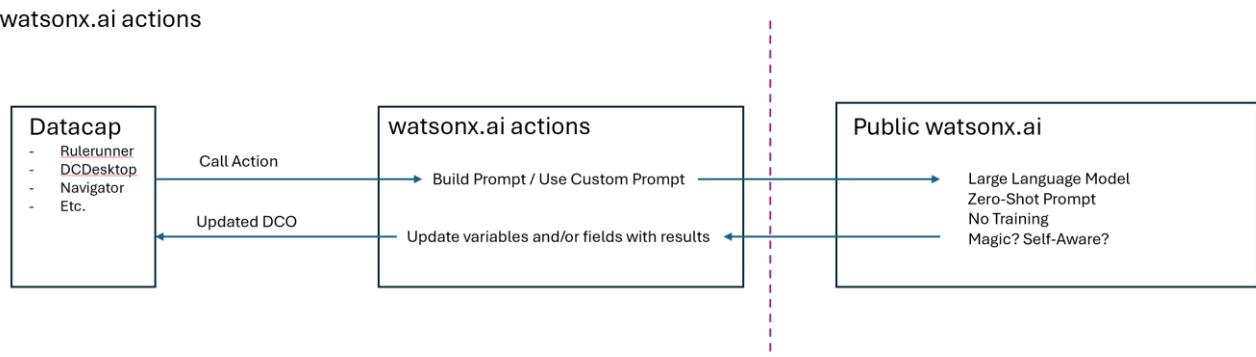
This lab demonstrates how to create a Datacap application and integrate it with *watsonx.ai* within IBM Cloud Pak for Business Automation. The focus is primarily on document capture and data processing capabilities using *Datacap in combination with watsonx.ai*. This case showcases how to classify documents and extract content using Watsonx.ai, simplifying the development, by reducing the need for complex rulesets such as fingerprinting, keyword or regex searches and ongoing maintenance. The Watsonx.ai integration also accelerates Datacap application development.

### How does it work?

### Datacap Document Processing



**Figure 1: Datacap Document Processing Functions**



**Figure 2: Watsonx.ai actions Flow diagram**

### 1.3 Overview of Exercises

#### Exercise 1:

In this exercise you will use the *FastDoc Admin authoring tool* to create an initial version of the DataCap application. FastDoc Admin lets you create applications quickly; however, it does not give you easy access to the full functionality of the Datacap product.

#### Exercise 2:

In this exercise, you will use the *Datacap Studio authoring tool* to enhance the newly created application by adding rulesets, rules, functions, task workflows and additional fields, allowing for capturing additional input forms and configuring communication to the ECM system. Datacap Studio is the more complex, but all-encompassing developer tool for Datacap applications.

#### Exercise 3:

This exercise guides you with running the Datacap application you developed in an environment that closely simulates that of a production scenario.

You will first use the *Datacap Navigator Desktop*, which acts as the runtime environment for your application. It allows you to capture input in batches, process documents, and export the extracted data and images to the ECM repository. This experience mirrors how virtual scanning would typically be initiated from the Content Navigator desktop.

Following that, you will use *IBM Content Navigator* as both the user interface and runtime environment. This demonstrates how Datacap can be accessed and operated entirely through a web browser. It also reflects real-world usage, where in a production setup, most tasks—such as classification, recognition, and export—are automatically handled by Rulerunner. In such cases, the human operator is generally responsible only for performing the ‘Verify’ task.

By completing these combined exercises, you will gain insight into both runtime options, desktop and browser-based and understand how tasks are distributed in an automated Datacap workflow.

#### Exercise Format:

- i) *Exercise Introduction*
  - Provides a high-level overview of the goals and objectives of the exercise.
  - Lists all required data, configurations, or prerequisites.
  - Designed so that experienced users can complete the exercise using this section alone.
- ii) *Step-by-Step Instructions*
  - Detailed, easy-to-follow instructions to guide users through the exercise.
  - Ideal for less experienced users or those unfamiliar with the tools or technologies involved.

- Each step is clearly described to ensure successful completion.
- iii) *Verification Instructions*
- Explains how to test and validate your implementation.
  - Helps ensure your configurations or development steps are functioning as expected before moving on.
- iv) *Exercise Summary*
- Recaps what has been completed in the exercise.
  - Reinforces key learning outcomes and successful configurations.

In the lab instructions there are sections marked with a grey bar on the left side, like this one. In these sections, additional explanations are given. For example, about concepts and additional features.

## 1.4 Audience

This lab is suited to anyone interested in the Digital Business Automation space, especially technical people who are interested in the details of processing input originating from paper forms, extracting image content to strings and numbers that will be fed into the other components like *IBM Business Automation Workflow* and processed there.

## 1.5 Getting HELP during the lab

You may reach out to the following resources at hand for any help needed during performing the lab:

- For **internal IBMers**, check the archived slack channel for questions: **ba-wxo-tech-jam-Datacap**. You may also use the Microsoft teams to answer the query.
- For **external participants** besides the Slack channel, use the *Microsoft Teams* to raise your query if you are in a TechJam event team meeting.
- Getting help after lab, reach out to the following:
  - [Sundeep.anne@ibm.com](mailto:Sundeep.anne@ibm.com)(preferred)
  - [harrish.narayanan@ibm.com](mailto:harrish.narayanan@ibm.com)
  - [pooja.luthra@ibm.com](mailto:pooja.luthra@ibm.com)

## 2. Lab Setup

### 2.1 Provision Techzone Environment

#### IBM TechZone – Overview

What is IBM TechZone?

IBM Technology Zone ([techzone.ibm.com](https://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.

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



**Note:** If you have been assigned a pre-provisioned DBA vM environment as part of the IBM event, you can skip this section, ‘2.1.1 Reserve Environment’, and continue to section ‘2.2’.

#### 2.1.1 Reserve Environment

- \_1. Navigate to:  
<https://techzone.ibm.com/collection/ibm-business-automation-traditional-and-on-premise/environments>.
- \_2. Click IBM Business Automation-Traditional and on-premises.v4.5 tile (at the right most end of the screen) and click on **Reserve Environments**

The screenshot shows a list of environment tiles. The first tile is for 'IBM Business Automation - Traditional and On-premises. V3.3 [Updated 2025-01-03]'. The second tile is for 'IBM Business Automation - Traditional and On-premises. V4.5 [Updated 2025-08-26] - US East...'. Both tiles show a brief description, update date, visibility, and a 'Reserve Environments' button at the bottom. A hand cursor is hovering over the 'Reserve Environments' button of the second tile.

- \_3. On Create a reservation screen **select option** for 'Education' in **Purpose**. which will give you 2 days to complete your lab exercises. Otherwise, you will need a legitimate opportunity to leverage another

reservation type.

The screenshot shows the 'Create a reservation' page in the IBM Technology Zone. The main form includes fields for Name, Purpose (with 'Education' selected), Sales Opportunity number, Purpose description, Preferred Geography, Start date and time, End date and time, and a checkbox for agreeing to terms and conditions. The right sidebar displays product details for 'IBM Business Automation - Traditional and On-Premises' and a 'Duration Policy' dropdown.

- \_4. Enter <some description> in the Purpose description box.
- \_5. For Preferred Geography (required) select your preferred data center location

The screenshot shows the 'Preferred Geography' section of the reservation form. A dropdown menu is open, displaying the selected value: 'itzvmware-itzna3-wdc06 - AMERICAS - us-east region - wdc06 datacenter'.

- \_6. Select a date a time if you picked something different then “Education” for the reservation. (if not, it will already be chosen for you)
- \_7. On the right-hand pane, click on the option “I agree” and click on the “Submit” button.

- \_8. You will receive an email saying ‘Reservation Provisioning on IBM Technology Zone’. Once provisioning gets completed, you will receive another email telling that your environment is ready. Once the start-up process is completed, 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](#).

 In the emails, pay attention to the Step and Status. The system is not up fully until the Step labeled “Cloud Pak Services Running” has a status of “Complete”.



If after receiving email, a few hours have passed but your environment is not up, check with Techzone support team via [How to Open a Support Web Case for IBMers](#).

- \_9. You will also need to reserve the **Watsonx.ai Techzone Tile** to obtain the Project ID and API key, which are required for integrating Datacap with Watsonx.ai Foundation Models.

Watsonx.ai Techzone Reservation Tile:

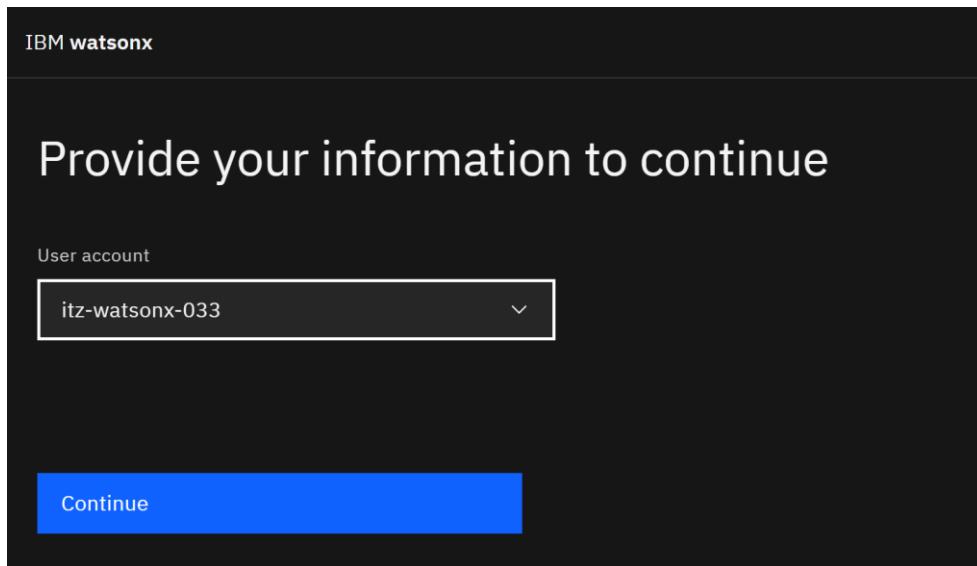
<https://techzone.ibm.com/my/reservations/create/6696a8ada7b42e001ef21b52>

To reserve this Techzone Tile, you must have a valid opportunity code.

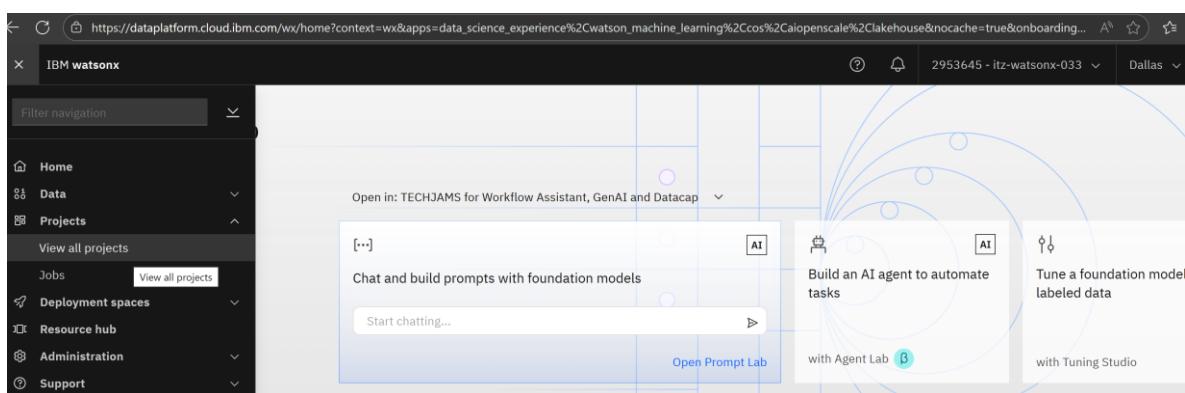
**Note:** Once you reserve the Watsonx.ai TechZone environment, you’ll receive multiple emails such as “Reservation Provisioning on IBM Technology Zone” and “Reservation Transferred on IBM Technology Zone.” If you don’t receive an invitation email to join the IBM Cloud account, you’ll need to use the AskTZ self-help chat. Provide the email ID you used to log into TechZone and the reservation ID from the confirmation email.

If your reservation is successfully retrieved based on the information provided, you’ll receive a join invitation email. Once you click “Join,” your ID will be mapped to the IBM Cloud account, giving you access to Watsonx Studio. From there, you can create a project and use its details—such as Project ID and API Key—for Datacap integration.

- \_10. Once you've joined the account, you can access watsonx.ai by clicking the provided link, selecting the account listed in your invitation email, and then clicking "Continue."



- \_11. Once you've successfully accessed the environment, you'll see the Open Prompt Lab. Click the hamburger menu, then select "View All Projects" to proceed.



- \_12. Click on "New Project," enter a name for your project, and then click the "Create" button to proceed.

A screenshot of a "Create a project" dialog box. It has tabs for "New", "Local file", and "Sample". The "New" tab is selected. The "Name" field contains "Datacap+Watsonx.ai". There are optional fields for "Description (optional)" and "Tags (optional)". The "Storage" section shows "itz-watsonx-033" and a note about Cloud Object Storage integration. At the bottom are "Cancel" and "Create" buttons, with "Create" being highlighted.

\_13. You'll see four tabs: Overview, Assets, Jobs, and Manage. Please click on the Manage tab to continue.

The screenshot shows the IBM Watsonx interface for a project named "Datacap+Watsonx.ai". The top navigation bar includes a menu icon, the "IBM Watsonx" logo, user information (2953645 - itz-watsonx-033), location (Dallas), and a "Launch IDE" button. Below the header, there are four tabs: "Overview" (selected), "Assets", "Jobs", and "Manage". A "Start working" section contains four cards with icons and text: "Add users as collaborators", "Add data to work with", "Chat and build prompts with foundation models", and "Tune a foundation model with labeled data". A "View all" link is located below this section. The main content area is divided into three columns: "Jump back in" (with a "View all" link), "Resource usage" (showing 0 CUH, 0 Tokens, 0 Hosting hours, and 0 Pages), and "Your documentation" (with a "New!" badge, a "Get started with your documentation" link, and a "Open Documentation editor" button). A "Project history" section shows a recent creation entry: "You created project Datacap+Watsonx.ai Today at 11:55 AM".

\_14. You can find the Project ID—needed for integrating Watsonx.ai with Datacap—under the Manage tab. Be sure to copy this ID for use during the configuration process.

The screenshot shows the "Manage" tab selected in the IBM Watsonx project settings. On the left, a sidebar menu includes "Project" (selected), "General" (highlighted with a grey background), "Access control", "Environments", "Resource usage", "Services & integrations", "Tools", and "Pipeline". The main content area is titled "General" and contains sections for "Details", "Description", "Tags", and "Project ID". The "Details" section shows the "Name" as "Datacap+Watsonx.ai". The "Description" section asks, "What's the purpose of this project?". The "Tags" section provides a placeholder: "Add tags to make projects easier to find.". The "Project ID" section shows a long, partially obscured string of characters.

\_15. From the left-hand menu, select Services and Integrations, then click on Associate Service to proceed.

- \_16. From the hamburger menu, click on IAM under the Administration section. This will open a new tab where you can create API keys. These keys will be valid for your IBM Cloud account and can be used across any of your projects.

- \_17. Click Create, enter a name for the key to help you remember its purpose, then copy and download the API key for safekeeping.

The screenshot shows the IBM Cloud IAM API keys management interface. The left sidebar has a 'API keys' section selected. The main area displays a table of API keys with columns for Status, Name, Description, Date created, and Enabled. Two entries are listed:

Status	Name	Description	Date created	Enabled
Open	TechJAm		10-1-2025 01:11 GMT	<input checked="" type="checkbox"/> Yes
Open	TigerTeamAPIKey		10-1-2025 01:12 GMT	<input checked="" type="checkbox"/> Yes

- \_18. Once you have the API key, return to the previous browser tab where Watsonx Studio is open. From the hamburger menu, click Home. You'll see the Prompt Lab featuring your newly created project. Click on the Freeform tab, where you can test your prompt and view the available foundational models, as shown in the screenshot.

The screenshot shows the Watsonx Prompt Lab interface. The top navigation bar includes 'Projects / Datacap+Watsonx.ai / Prompt Lab'. The main area has tabs for Chat, Structured, and Freeform, with Freeform selected. A sidebar on the left shows 'Sample prompts' with a 'Try these prompt tips' button. The central workspace has a 'Hint' message: 'Hint: This model works better when you provide at least 1 example.' Below it, there's a text input field with placeholder text: 'Enter your prompt text. Remember: This is not a chat interface. Provide instructions and examples to show the model what to do.' At the bottom, there's a note: 'When you prompt a text-generating model, the model responds by appending text to your prompt text or continuing your prompt text.' A right-hand sidebar shows 'Model: granite-3-8b-instruct' and a 'View all foundation models' button.

- \_19. Switch the model to “mistralai/mistral-small-3-1-24b-instruct-2503” as it was used for this particular use case. You’re also free to explore other available models if needed.
- \_20. To retrieve the necessary integration details—Project ID, Endpoint URL, and Model Name—click on the View Code tab located at the top-right corner next to the foundational models. Copy the displayed cURL command into a notepad; it contains all the required parameters for integrating Datacap with Watsonx.ai.

The screenshot shows the Prompt Lab interface with the following details:

- Projects / Datacap+Watsonx.ai / Prompt Lab**
- AI guardrails on**
- Model: mistral-small-3-1-24b-instruct-2503**
- Chat, Structured, Freeform** tabs
- AI** button
- View code** section with **Curl, Node.js, Python** options
- Hint:** This model works better when you provide at least 1 example.
- Enter your prompt text.**
- Remember:** This is not a chat interface. Provide instructions and examples to show the model what to do.
- When you prompt a text-generating model, the model responds by appending text to your prompt text or continuing your prompt text.**
- Try the sample prompts for a variety of use cases.**
- Summarization:** **Meeting transcript summary** (Summarize the discussion from a meeting transcript).
- Classification:** **Scenario classification** (Classify scenario based on project categories), **Feedback classification** (Classify feedback about insurance customer service).
- Generation:** **Claim processing generation** (Generate the next steps for).
- Code Sample (Curl):**

```
curl "https://us-south.ml.cloud.ibm.com/ml/v1/text/generate?version=2023-05-29" \
-H 'Content-Type: application/json' \
-H 'Accept: application/json' \
-H "Authorization: Bearer ${YOUR_ACCESS_TOKEN}" \
-d '{
  "input": "",
  "parameters": {
    "decoding_method": "greedy",
    "max_new_tokens": 200,
    "min_new_tokens": 0,
    "stop_sequences": [],
    "repetition_penalty": 1
  },
  "model_id": "mistralai/mistral-small-3-1-24b-instruct-2503",
  "project_id": "soc3d58e357b"
}'
```

### Note:

- Reservations are restricted to Demo or Pilot use only.
- Education and testing reservations are currently unavailable due to high operational costs in this environment.

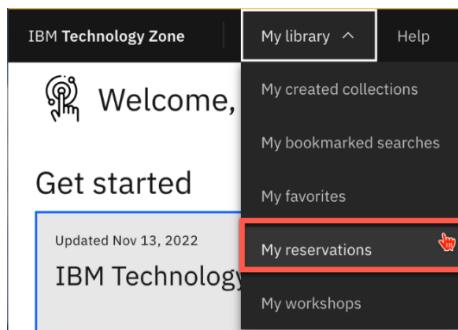
For **Tech Jam events**, a *temporary project ID and IAM key* with limited tokens will be provided by the event organizers.

Thank you for your understanding and cooperation.

- \_21. Once you get the email from the IBM Technology Zone site, you can access your environment reservation(s) by **clicking** on the View My Reservations link in email.



Or you may access the same from your reservations in the Techzone environment under **My library** then clicking on **My Reservations**



You can also access directly using the link below

<https://techzone.ibm.com/my/reservations>

## 2.2 Getting Started – Accessing the Environment

This section explains how to access the DBA VM environment prepared for DBA Demos and Labs v25.0.0, and how to get familiar with the key tools required for the upcoming exercises.

Tool	Location / URL
Datacap FastDoc	<b>Start -&gt; Datacap FastDoc (Admin)</b>
Datacap Studio	<b>Start -&gt; Datacap Studio</b>

For this, you will require the following IDs and passwords:

Access Mode	User ID	Password
Windows Console	<b>Administrator</b>	<b>IBMDem0s!</b>
Datacap FastDoc/Desktop	<b>admin</b>	<b>admin</b>

Step 1: Open the reserved environment by navigating to the TechZone

URL:<https://techzone.ibm.com/my/reservations>

Step 2:

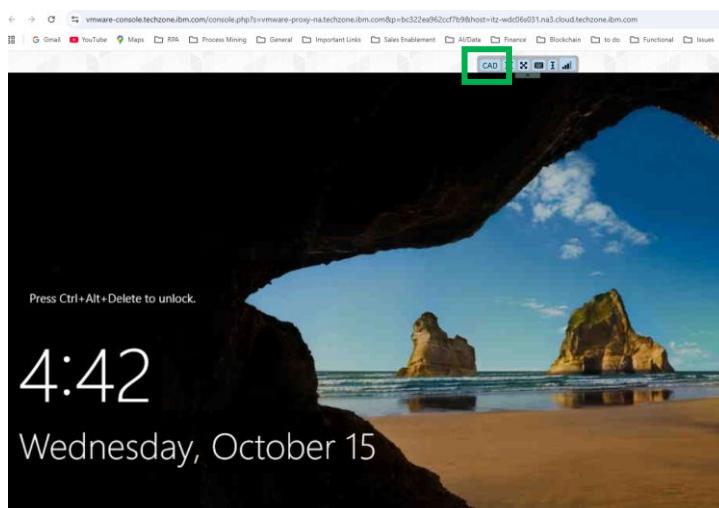
Click on the **Console** icon from the **Status** page, as shown below.

The screenshot shows the 'Reservation Status' page for a specific reservation. It includes sections for 'Published services', 'Purpose', 'Environment', and 'Virtual Machines'. A red arrow points to the 'Published services' section, highlighting the 'Remote Desktop (RDP)' link. Another red arrow points to the 'Virtual Machines' section, highlighting the 'Console' button for the first listed VM.

### Step 3:

The console window opens, if the windows appears in locked state, unlock the windows using the controls at the top of console(CAD) and enter the following credentials to login:

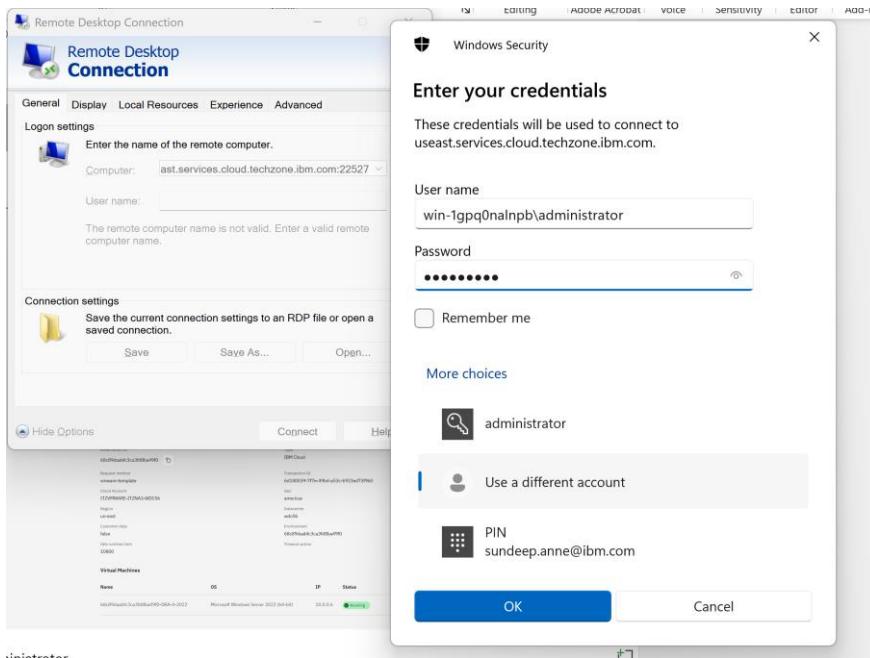
**Username** – Administrator  
**Password** – IBMDem0s!



**Alternatively**, you can connect to the virtual machine using the **Remote Desktop Client** on your local machine.

To do this, copy the **RDP** link from the **Reservation Status** page and paste it into your RDP client. Use the following credentials:

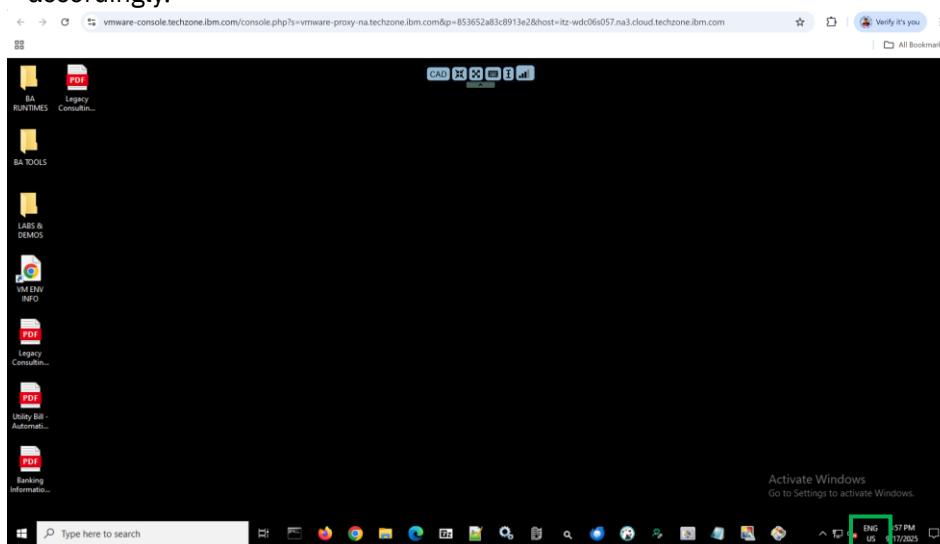
- **Username:** win-1gpq0nalnpb\administrator
- **Password:** IBMDem0s!



**Step 4:** The VM's desktop will open in a new browser tab.

### 2.3 Introducing the components

- As a first step, you may want to change the keyboard layout from **US English** to your preferred language. You can do this at two locations using your mouse:
  - In the **Windows taskbar** (bottom right corner) – click the language/icon to switch layouts.
  - In the **browser's remote desktop toolbar** (if available) – adjust the keyboard input settings accordingly.



- Have a look at the start menu. Here you find:

- ✓  **Datacap FastDoc**, one of the two development environments. It will be used in Exercise 01 to build a first version of the application.



✓ **Datacap Studio**, the other development environment which will be used in Exercise 2 to enrich the application with further fields, features and functionality.

- ✓ To use **Datacap Navigator** to run the Datacap “**ClientDocumentProcessing**” application which will be used in Exercise 3. ‘<https://win-1gpq0nalnpb:9443/navigator/?desktop=ClientDocumentProcessing>’

A screenshot of a web browser window displaying the Datacap Navigator interface. The title bar shows the URL: "win-1gpq0nalnpb:9443/navigator/?desktop=ClientDocumentProcessing". The main content area is titled "IBM Capture Demonstration" and "ClientDocumentProcessing". It features a central table with columns: Queue, Batch, Job, Task, Status, Job Start, Job Time, Operator, Station, Docum..., and Pages. A single row is visible: "20250916.000007" under Queue, "Navigator Job" under Job, "Export" under Task, "Job done" under Status, "9/26/2023, 9:54 PM" under Job Start, "1" under Job Time, "admin" under Operator, "1" under Station, "1" under Docum..., and "1" under Pages. To the left of the table is a sidebar with icons for Scan, Verify, and Upload, along with a "Saved Filter" section. To the right of the table is a "Properties" panel showing "Class: Queue" and "Batch: 20250916.000007".

### 3. Exercises – Configuration Lab Instructions

#### 3.1 Exercise 1: Create Datacap application by using Datacap FastDoc Admin

##### 3.1.1 Datacap FastDoc – Introduction

In this exercise, you will explore the user interface of Datacap FastDoc.

Datacap FastDoc	Start -> Datacap FastDoc (Admin)
-----------------	----------------------------------

**End result/Conclusion:** The exercise is completed when you build the first version of the ClientDocumentsProcessing solution and extract the desired data fields from the Client Onboarding documents.

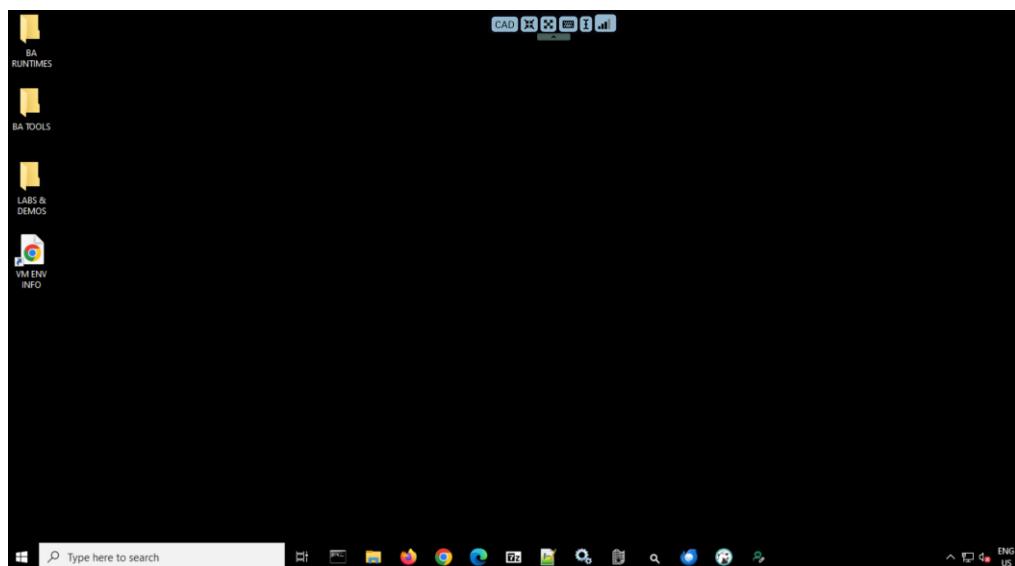
Using FastDoc, you'll learn that the overall processing of a Datacap application follows a more or less fixed sequence of major steps (the Datacap workflow).

**Note:** You use FastDoc Admin and Datacap Studio tools throughout the exercises.

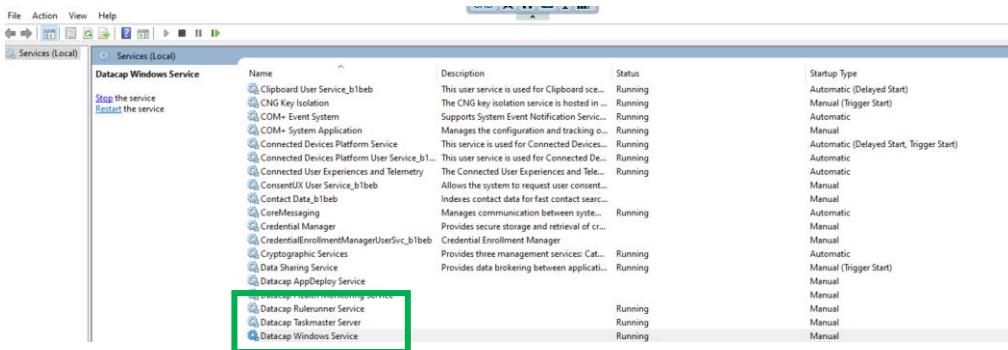
##### 3.1.2 Datacap FastDoc – Step by Step Instructions

To get started with this exercise, execute the following steps:

1. Log in to VM using the credentials as relevant provided in **section 2.2**.



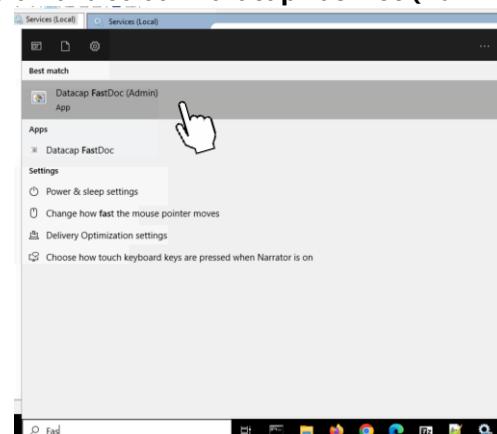
2. Ensure Datacap Server is started. Go to “Services” and check the status of three Datacap services as listed below:
  - ✓ Datacap Taskmaster Service
  - ✓ Datacap Rulerunner Service
  - ✓ Datacap Windows Service



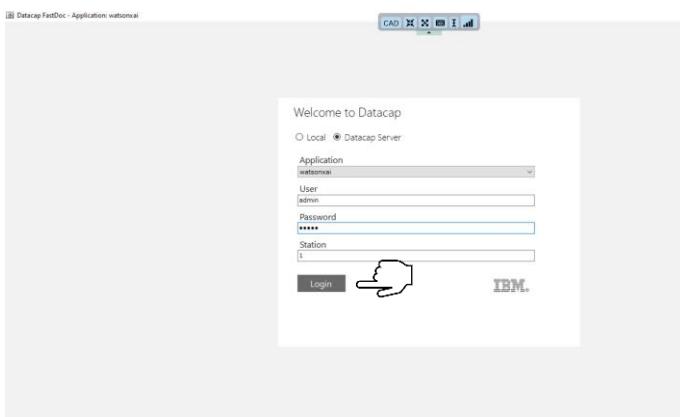
3. If it's not already running, start it by right clicking on the Datacap Task Master and do the same for the rest of the mandatory services enlisted in point 2.

The first screenshot shows the Windows Services (Local) window with the Datacap Taskmaster Server service selected. A blue box highlights the "Start" button. The second screenshot shows the service starting, with a blue box highlighting the progress bar. The third screenshot shows a "Service Control" dialog box with the message "Windows is attempting to start the following service on Local Computer..." and a "Close" button.

4. Click on Start and select '**Datacap FastDoc (Admin)**'.



5. Select "**Datacap Server**", choose any application, and log in using the following credentials:  
 • **Username:** admin, **Password:** admin and **Station:** 1

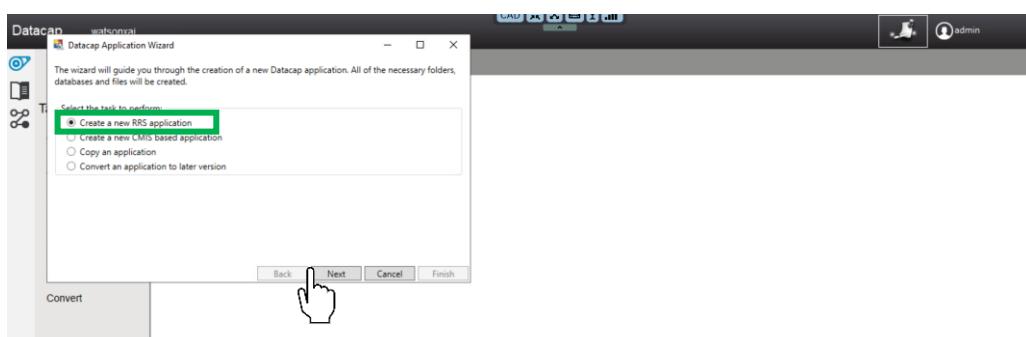


6. Explore the user interface of Datacap FastDoc:

- ✓ Explore the different sections (Top bar, leftmost column)
- ✓ Click on the different icons in the leftmost column. For each of these icons, the rest of the UI shows what you can do and what you can act upon.
- ✓ is for processing batches of documents  
You see various actions that you can perform on batches of documents.
- ✓ is for configuring documents, pages and fields. You can browse through the structure of a document with its pages and fields.
- ✓ is for configuring the Datacap workflow  
You see various types of jobs, the high-level steps of the job's workflow and which processing rules are associated with each of these steps.

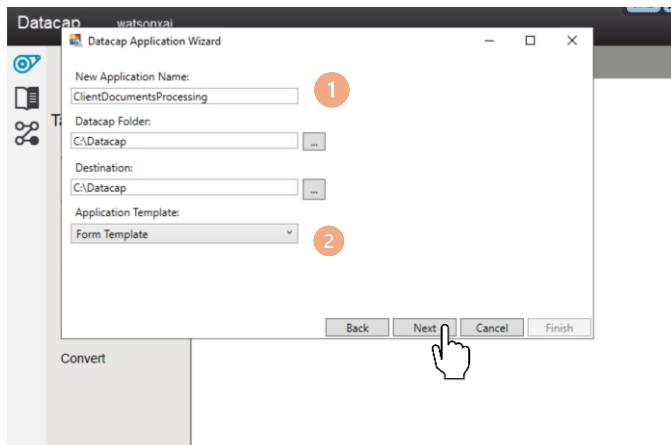
7. Now you will start to define the application for this exercise:

8. In the top right corner bar, click on Application Wizard ()
9. Select '*Create a new RRS application*' and click on 'Next' button.

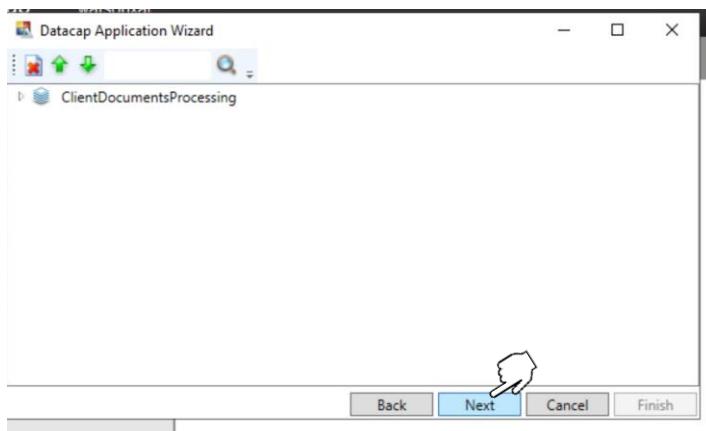


(RRS stands for RuleRunner Service, which is one of Datacap's internal components)

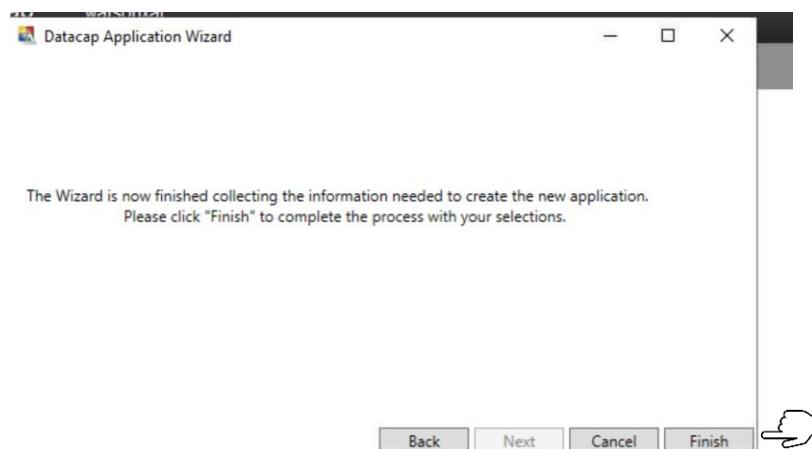
10. Enter in *New Application Name*: **ClientDocumentsProcessing** and set value of *Application Template*: **Form Template**. Click on **Next**.



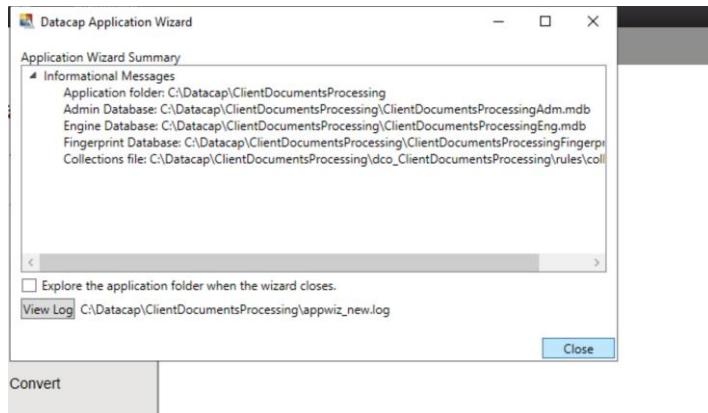
11. Client Documents Processing is displayed, click **Next**.



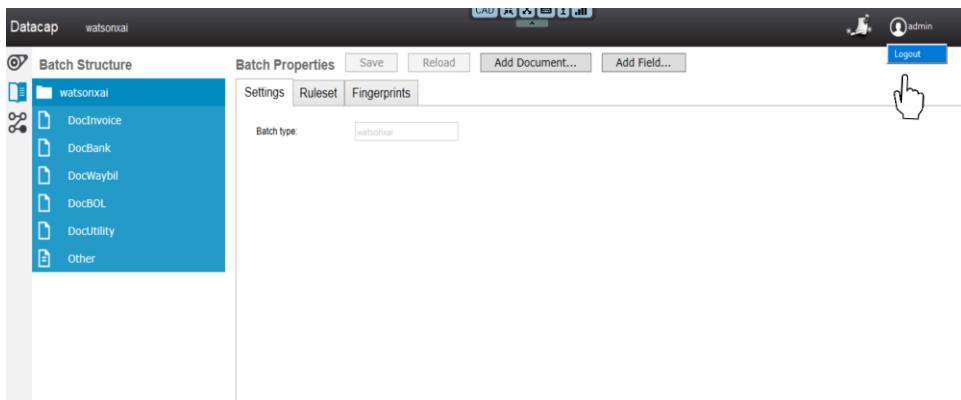
12. Click **Next** again in the next window. Click on **Finish** in the following window to finish the process of selections.



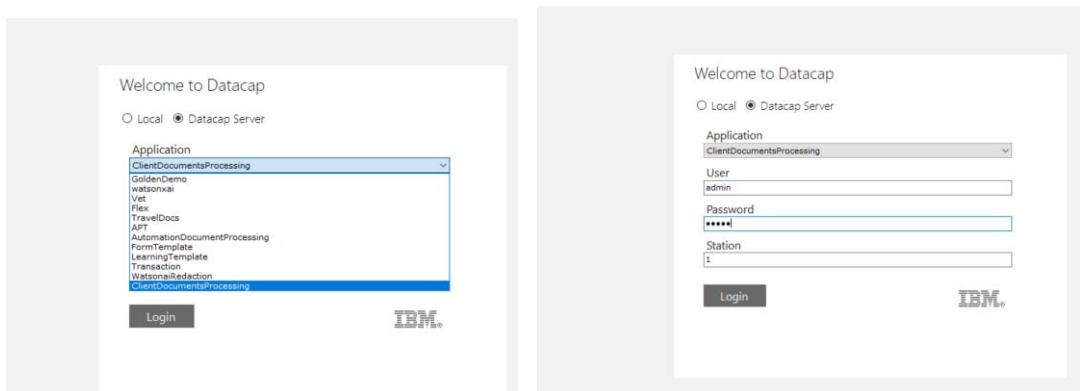
13. Once the Datacap application is successfully created, you will see the Datacap Application Summary screen as shown below. Click on **Close** button.



14. Log out of the Datacap FastDoc application .

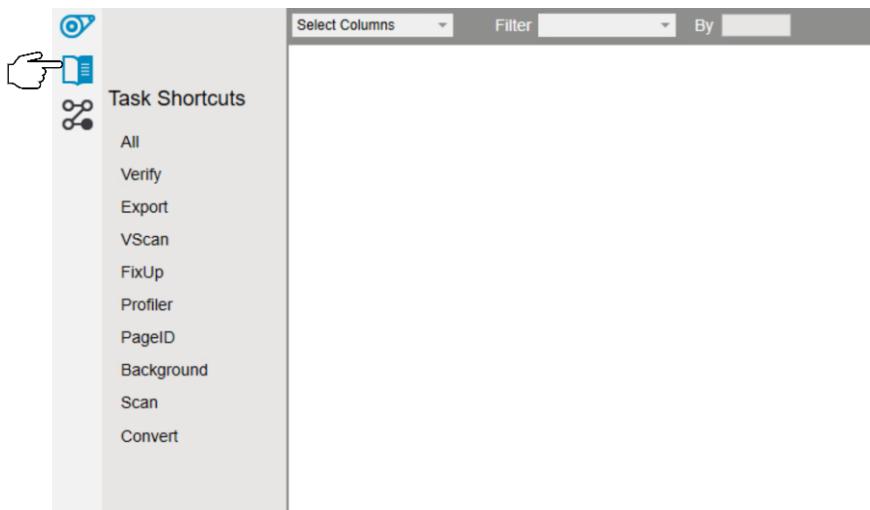


15. Log in again by first selecting **ClientDocumentsProcessing** application and using the **admin/admin** credentials with **station 1**. Click on **Login** button.

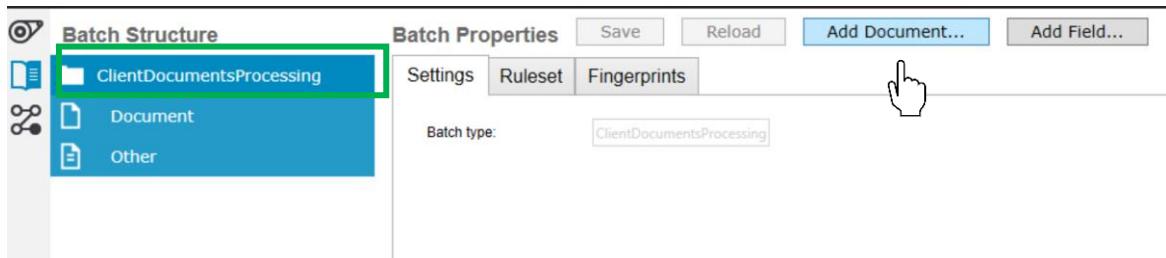


*You are creating the Client Document Type and Page Types using the FastDoc app.*

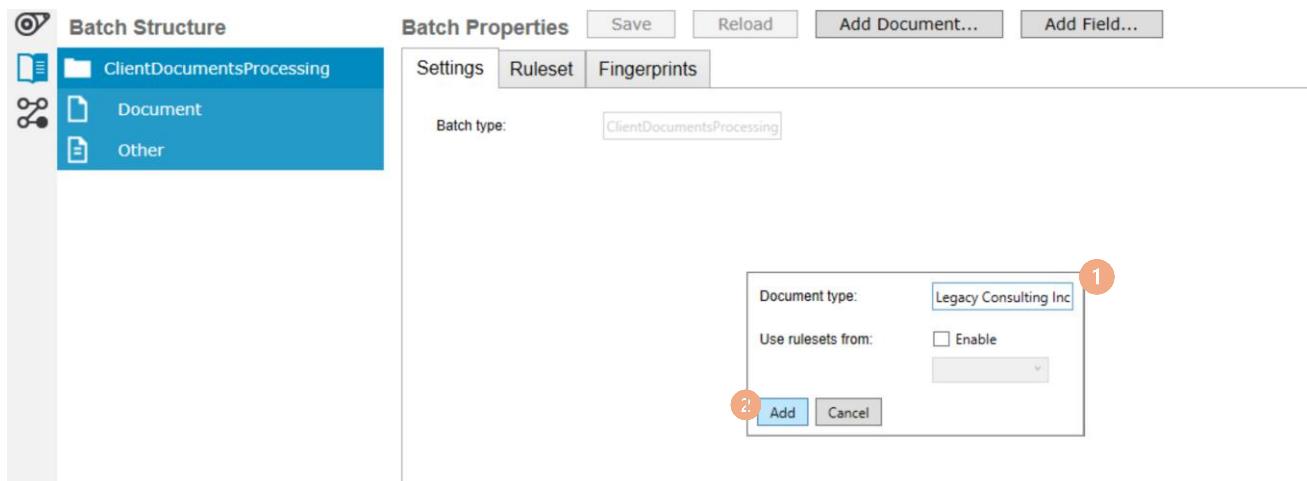
16. Click on the icon which is for configuring documents, pages and fields.



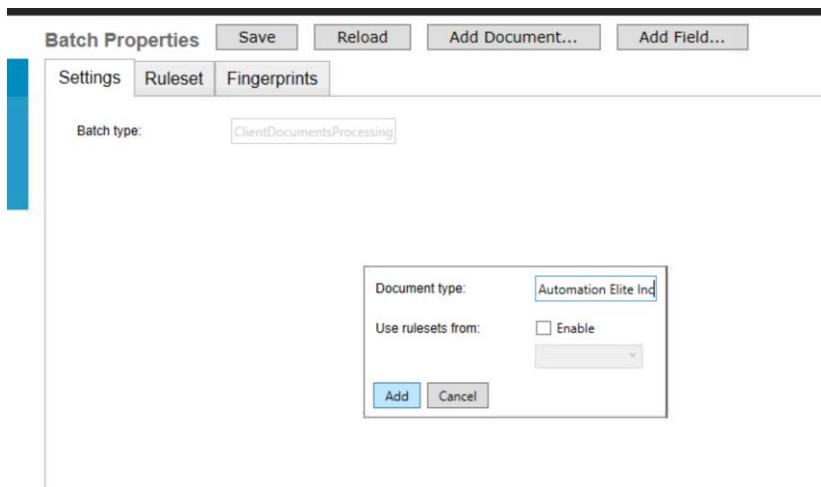
17. Select ClientDocumentsProcessing under Batch Structure and click on Add Document.



18. Name the document type as **Legacy Consulting Inc.** and click on **Add** button.

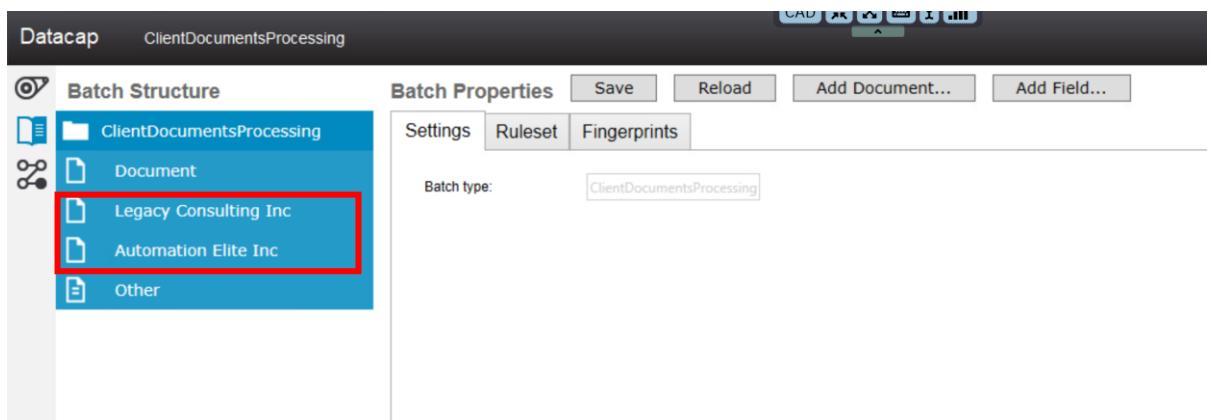


19. Click **Add Document** again to create another document type, and name it **Automation Elite Inc.**

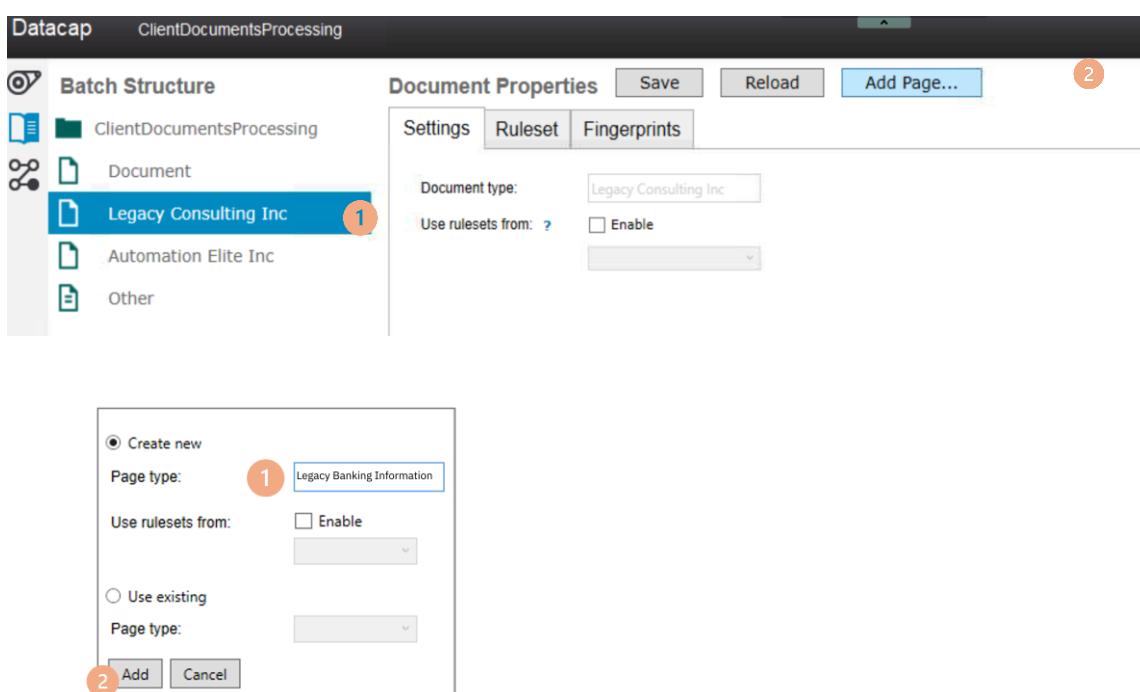


20. Click on Save button

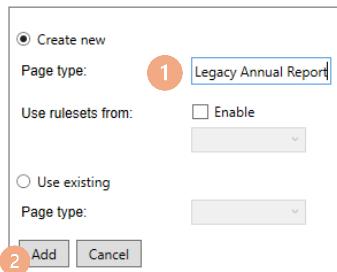
21. You can see two new document types created as shown in the screenshot below.



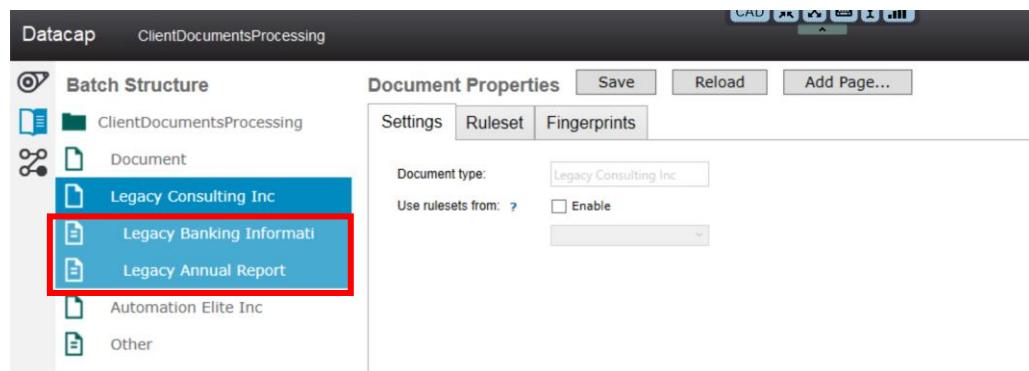
22. To add two new page types for the **Legacy Consulting Inc** document type, select **Legacy Consulting Inc** under **Batch Structure**, click **Add Page**, and name the first page type **Legacy Banking Information**.



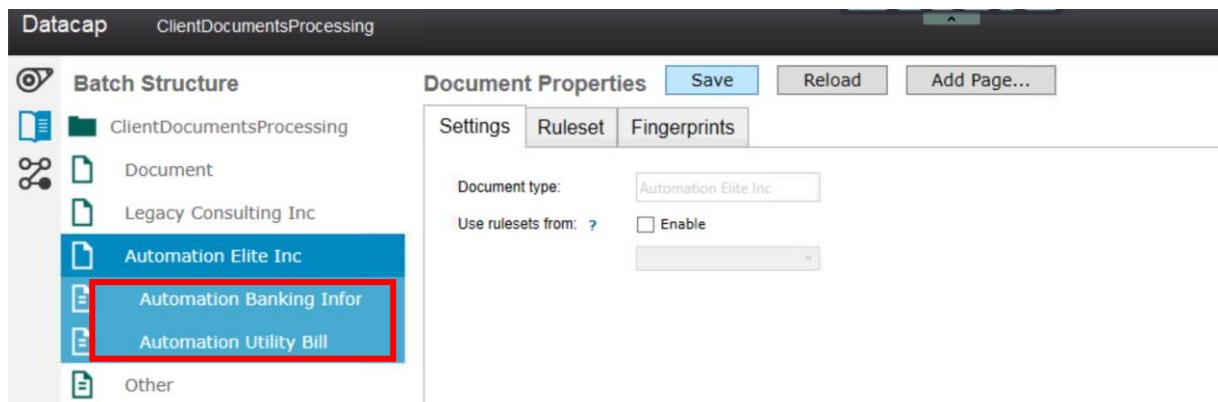
23. Click **Add Page** again and name it **Legacy Annual Report**. Click on **Add** button.



24. After adding these two-page types to the **Legacy Consulting Inc** document type, you should see the structure as shown in the screen below.



25. Add two page types—**Automation Banking Information** and **Automation Utility Bill** to the Automation Elite Inc document type.



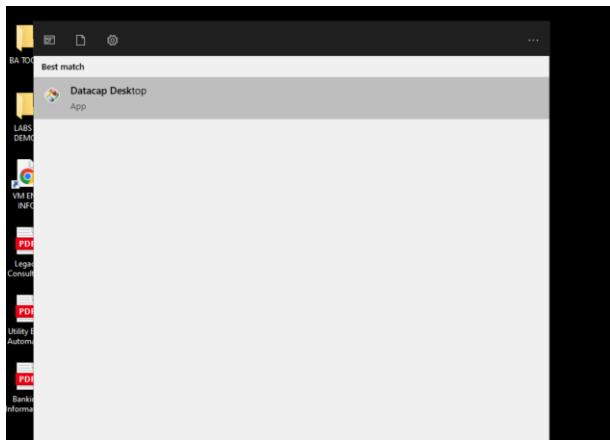
26. Click on Save button

### 3.1.3 Datacap FastDoc – Verification Instructions

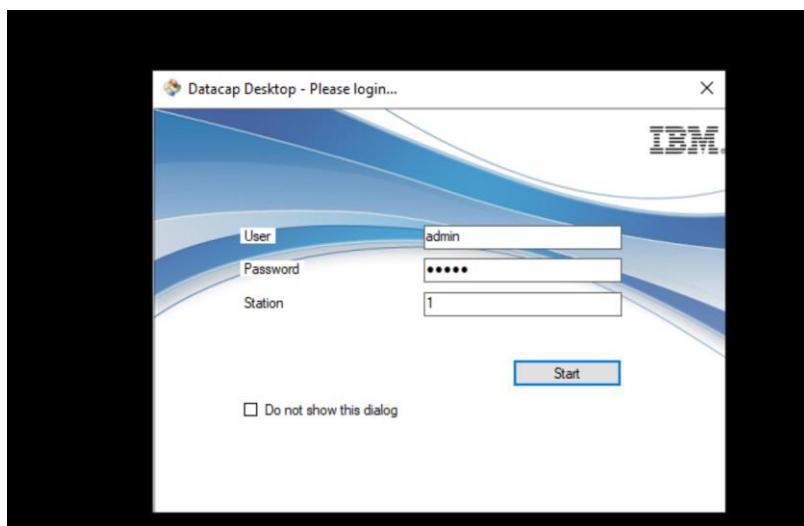
To verify successful completion of this exercise, ensure that you can run the newly created **ClientDocumentsProcessing** application.

\_1. Start Datacap Desktop:

Click on **Start** menu and select **Datacap Desktop**

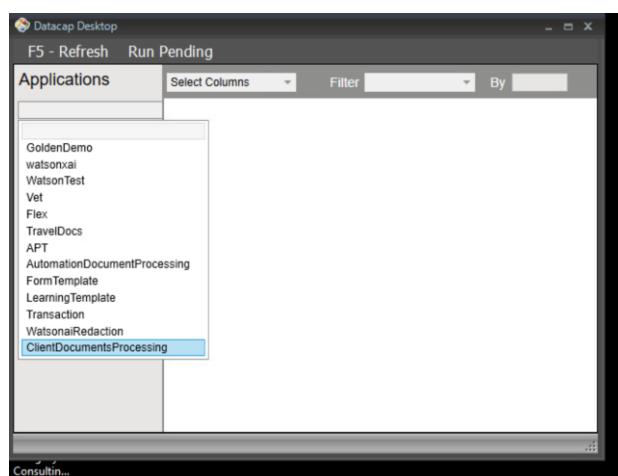


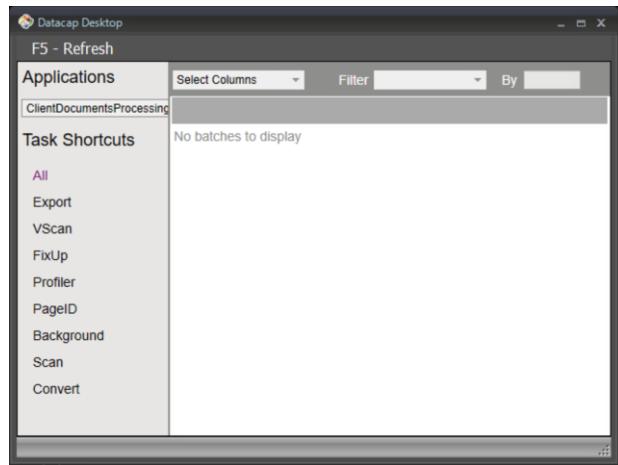
\_2. Log in as user – ‘admin’, password – ‘admin’ and station – ‘1’



\_3. What you see next is the **workplace** of the typical operator who uses physical scanners to process batches of documents through the **Datacap Thick Client Desktop**.<Explain more>.

\_4. Open the **ClientDocumentsProcessing** application in Datacap Desktop and you will see the application getting loaded.





\_5. Once the application loads successfully in Datacap Desktop, it confirms that the initial Datacap application has been created correctly. At this point, you can close both the FastDoc Admin and Datacap Desktop clients.

### 3.1.4 Datacap FastDoc – Summary

In this exercise, you have:

1. Gained an initial impression of Datacap FastDoc.
2. Explored the various tabs and sections within the application.
3. Developed a basic Datacap application, including creating the batch structure with document and page types.
4. Verified that the **ClientDocumentsProcessing** application runs successfully from the Thick Client.

**Note:** If the Datacap application is not created correctly or contains issues, you may encounter errors when attempting to open it in the Datacap Desktop application.

## 3.2 Exercise 2: Enhance and complete the Datacap application by using Datacap Studio

### 3.2.1 Datacap Studio

In this exercise, you will explore the user interface of **Datacap Studio**, the advanced development tool for Datacap.

You will access the following tools:

Tool	Location / URL
Datacap Studio	<a href="#">Start -&gt; Datacap Studio</a>

The exercise will be completed once you have successfully accessed all the Datacap development tools and completed the verification steps.

If you want to get a more detailed set of information on Datacap Studio, please follow the below link to the Datacap Knowledge Center describing the functions and concepts of Datacap Studio:

<https://www.ibm.com/docs/en/datacap/9.1.9?topic=reference-datacap-studio>

### 3.2.2 Datacap Studio – Step by Step Instructions

#### Section Overview: Enhancing and Completing the ClientDocumentsProcessing App in Datacap Studio

In this section, you will perform the following key actions to enhance and complete the **ClientDocumentsProcessing** application for document processing using Datacap Studio.

#### Document Classification using Watsonx.ai Large Language Models (LLM)

- Create Watsonx.ai\_PageID Ruleset(Rules, and Functions) for the ClientDocumentsProcessing app with Datacap Studio
- Map Watsonx.ai\_PageID Ruleset to Task Profile Workflow
- Map the Watsonx.ai\_PageID (Other Rule) to the Document Hierarchy (DCO)

#### Document Extraction using Watsonx.ai Large Language Models (LLM)

- Create Extraction with Watsonx.ai Rulesets(Rules, and Functions) for the ClientDocumentsProcessing app with Datacap Studio
- Map Extraction with Watsonx.ai Ruleset to Task Profile Workflow
- Map the Extraction with Watsonx.ai (Extraction Rule) to the Document Hierarchy (DCO)

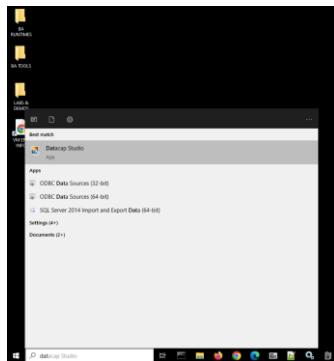
#### Document verification

- Create verify Ruleset(Rules, and Functions) for the ClientDocumentsProcessing app with Datacap Studio
- Map verify Ruleset to Task Profile Workflow

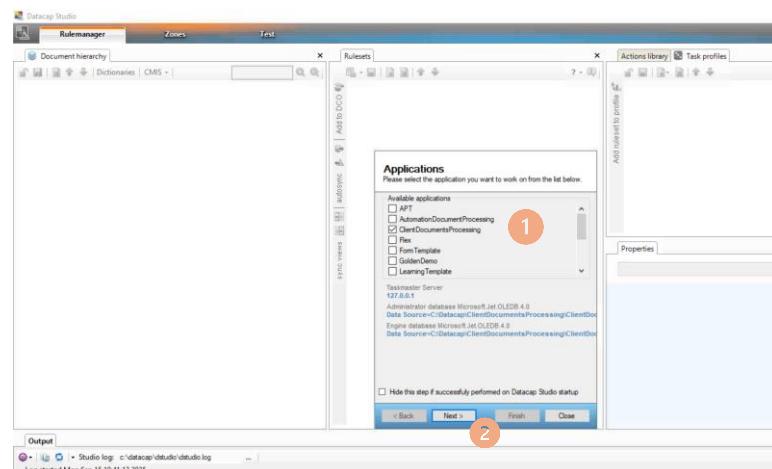
#### Configure FileNet Content Manager integration

- Set up **FileNet** to export extracted fields and documents from Datacap

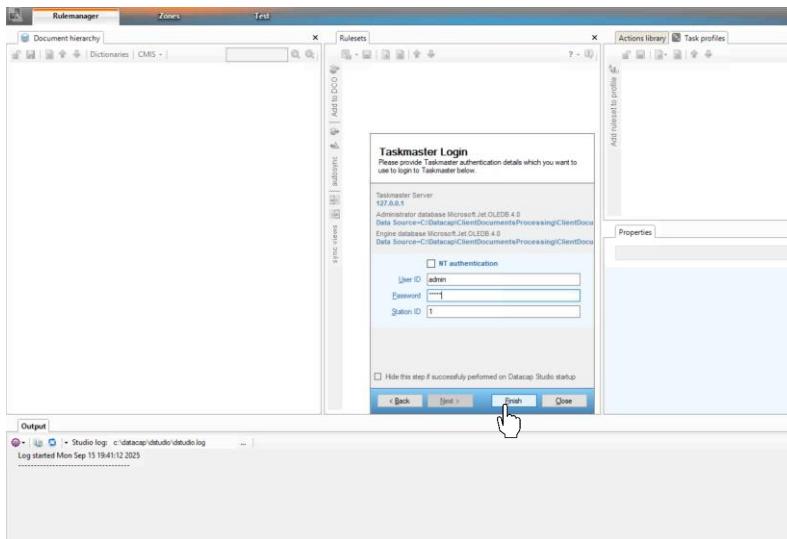
\_1. Click on Start and open Datacap Studio.



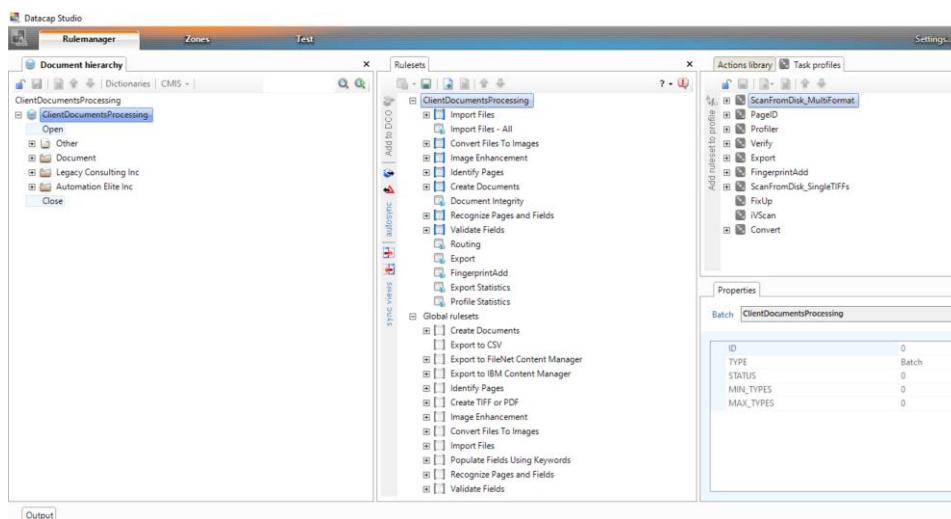
\_2. Select the application you created in exercise 2 and click Next.



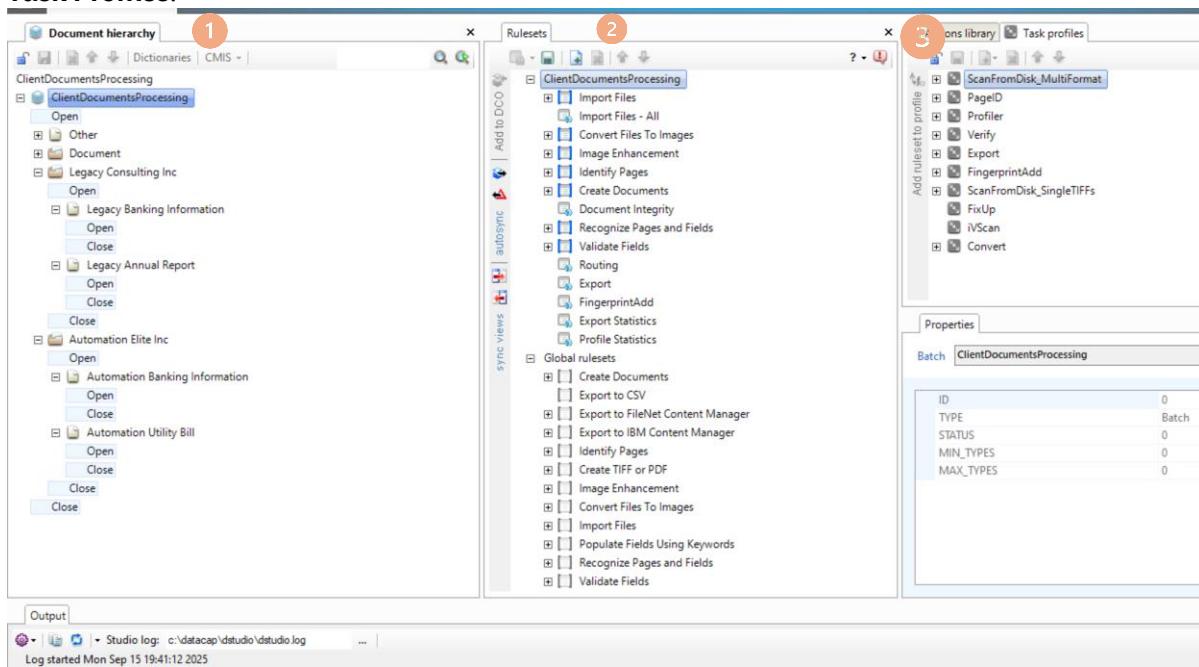
\_3. On the Taskmaster login window, enter the username **admin**, password **admin**, and set the station ID to 1. Click on Finish button to Login.



\_4. Explore the user interface of Datacap Studio.

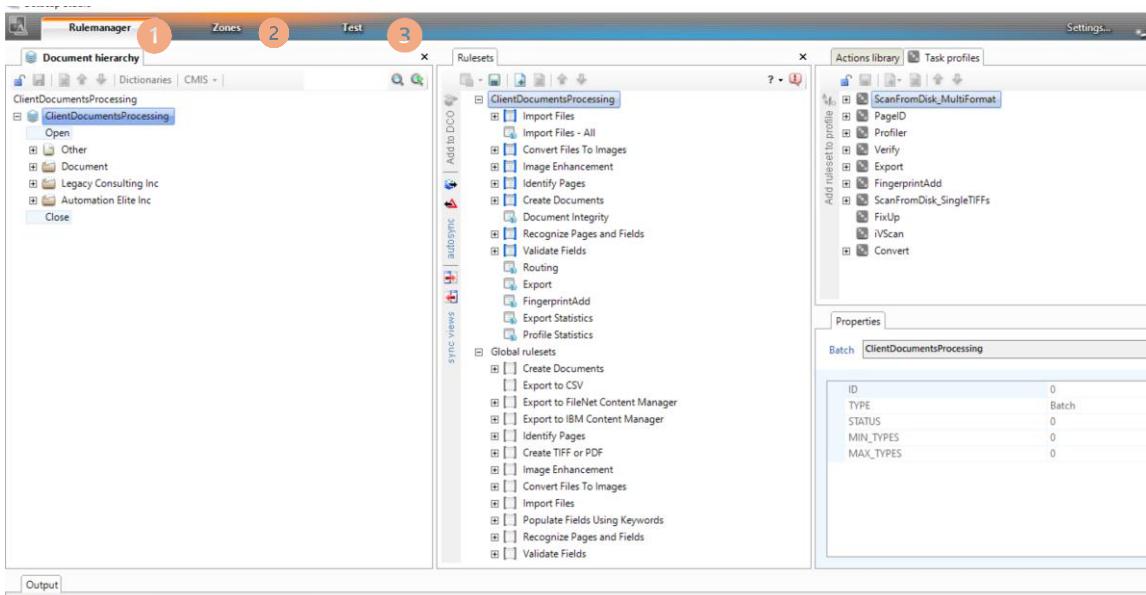


- a. Explore the different sections: **Document Hierarchy**, **Rulesets**, **Properties**, **Actions Library**, and **Task Profiles**.



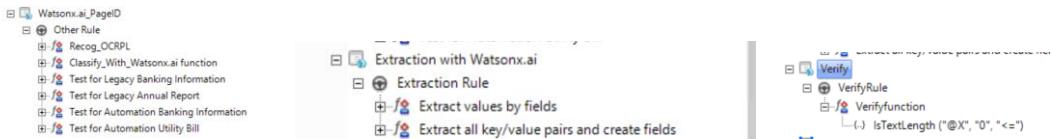
- b. Explore the different tabs from the top ribbon: **Rulemanager**, **Zones**, and **Test**.

Note: Some rulesets and task profile tasks may be created automatically when you select the **Form Template** application during setup.



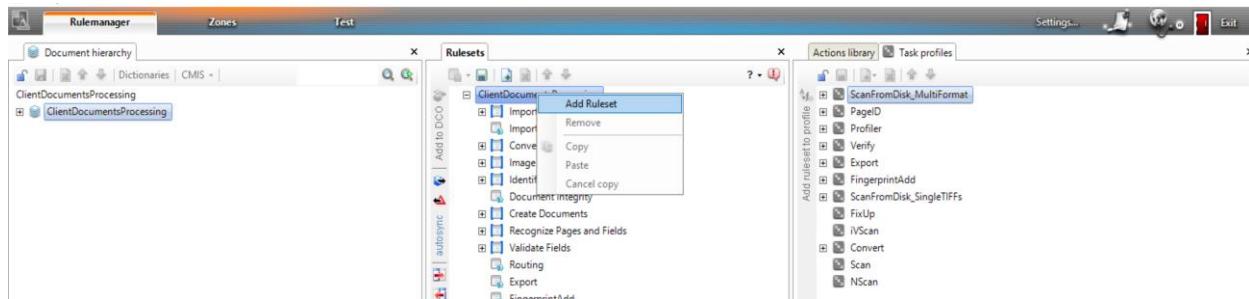
### 3.2.2.1 Document Classification using Watsonx.ai Large Language Models (LLM)

Create the Watsonx.ai\_PageID ruleset (including rules and functions) for the ClientDocumentsProcessing application using Datacap Studio. In this section, you will create the PageID ruleset along with the relevant rules and functions, as illustrated in the diagram below.



Key advantage of integrating Watsonx.ai with IBM Datacap for document processing: Enhanced semantic comprehension of unstructured data

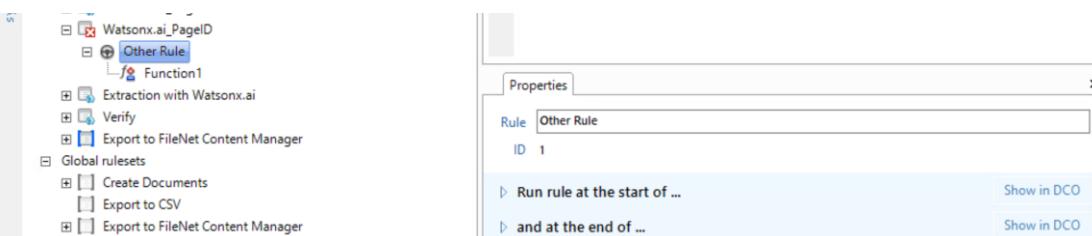
\_1. Right Click on the **ClientDocumentProcessing** under rulesets and click on ‘Add Ruleset’ to add new ruleset called “**watsonx.ai\_PageID**” to recognise the documents, pages and for Document classification. You can also click on the icon to add new ruleset by selecting ClientDocumentsProcessing application.



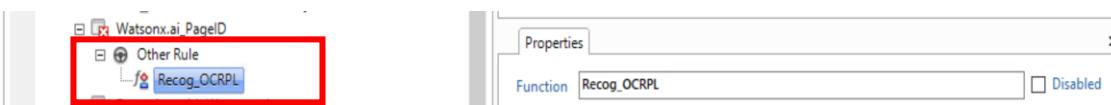
\_2. Name the ruleset “**Watsonx.ai\_PageID**”.



\_3. Right click on Watsonx.ai\_PageID ruleset and select Add Rule. Name the rule “**Other Rule**”, as shown below.

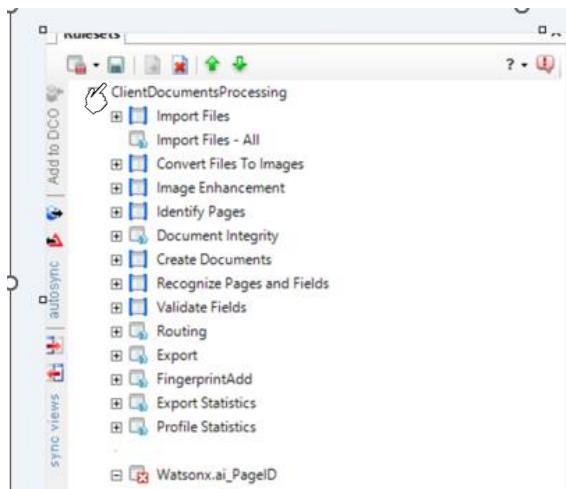


\_4. Right click on Other Rule and select Add Function to add a new function named **Recog\_OCRPL**.

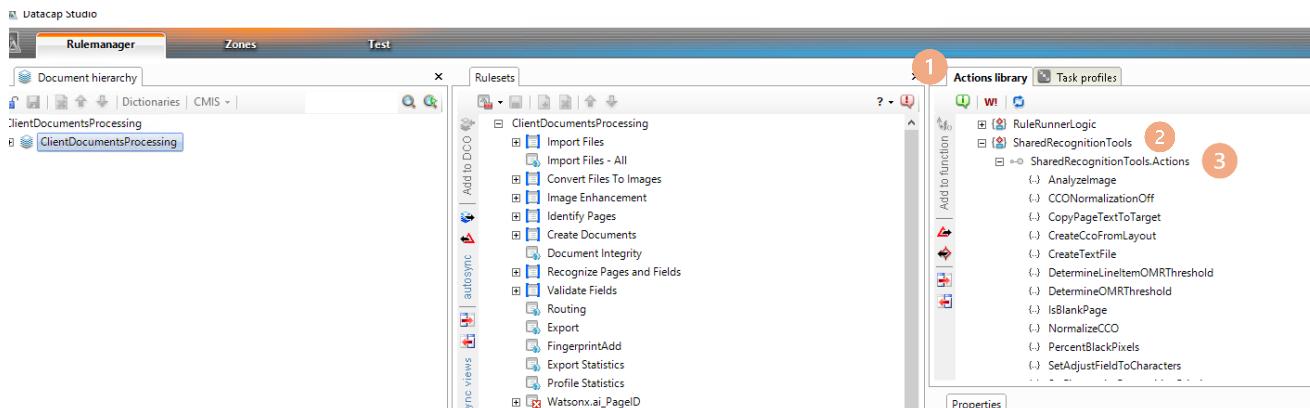


\_5. Click on **Save Button**.

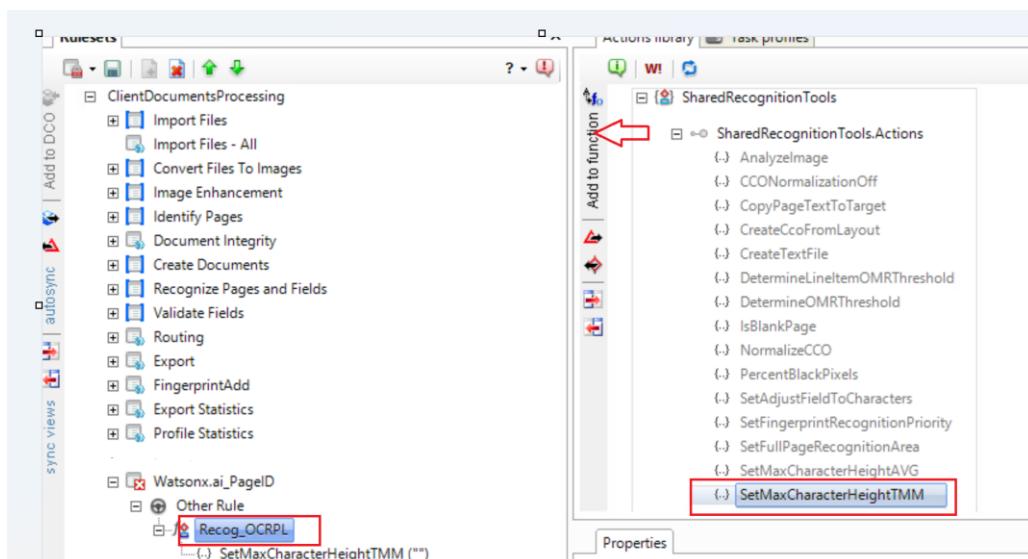




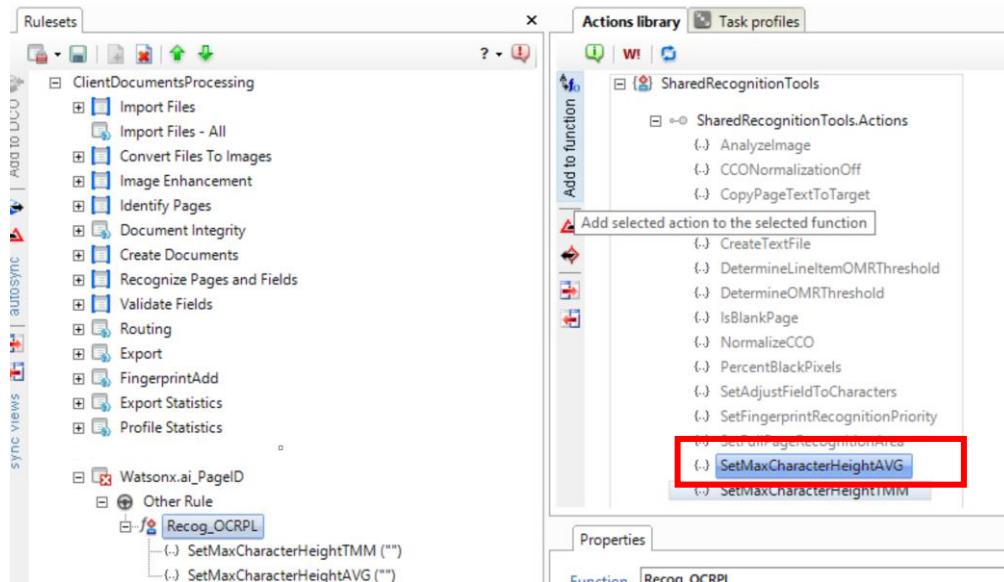
\_6. Go to actions library, scroll down to SharedRecognitionTools. Open actions under SharedRecognitionTools by clicking on SharedRecognitionTools.Actions.



\_7. Select the **SetMaxCharacterHeightTMM** method from the actions. On the Rulesets section, select Recog\_OCRPL function and click on 'Add to function' button from Action Library's panel.



\_8. Add the **SetMaxCharacterHeightAVG** method from the **Recognition Tools** actions available under the **SharedRecognition Tools** action library to the **Recog\_OCRPL** function.

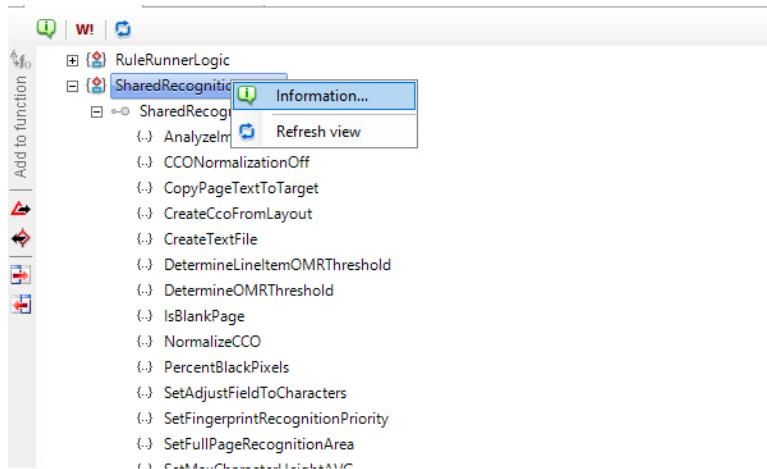


You can also right-click on any action within the Action Library in Datacap Studio to view detailed information and reference documentation about that action.

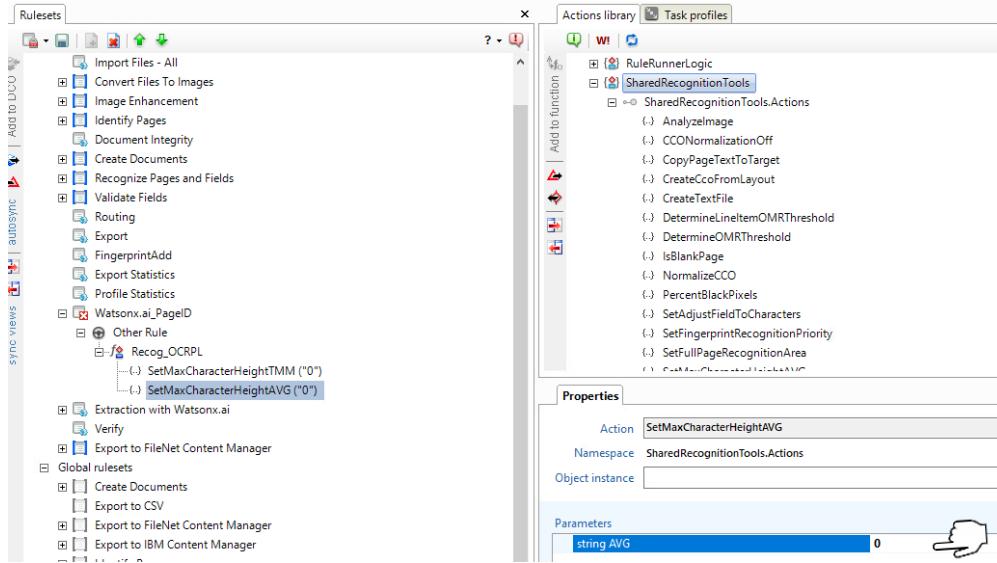
This feature provides helpful insights into:

What the action does

- Input and output parameters
- Usage guidelines and common scenarios



\_9. Select '**SetMaxCharacterHeightTMM**' from recently created in Ruleset panel, you will see **Properties** tab appear on the right hand side. Set the values of parameters of both **SetMaxCharacterHeightTMM** and **SetMaxCharacterHeightAVG** methods to **0**.



\_10. Open **RecognitionOCRPL** and open Actions underneath that. From there choose '**Recognize**' and add it to **Recog\_OCRPL function** as done earlier.

Datacap supports multiple embedded OCR engines beyond the default configurations. These include:

ABBY OCR (OCRA)

Nuance OCR (OCRS/R)

You are encouraged to experiment with these alternative recognition engines to understand their strengths, limitations, and use cases. Each engine may offer varying performance based on:

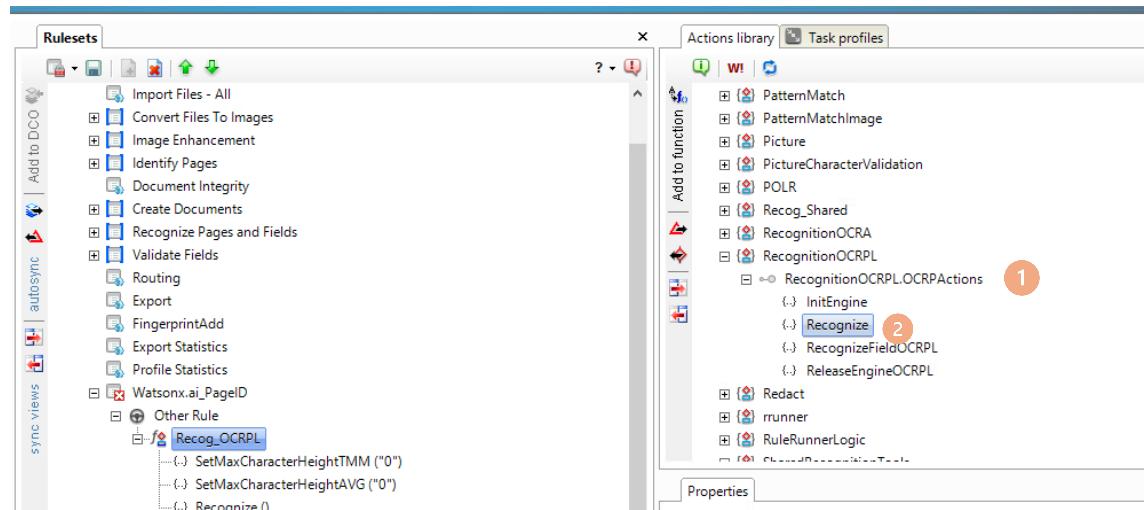
Document layout complexity

Language support

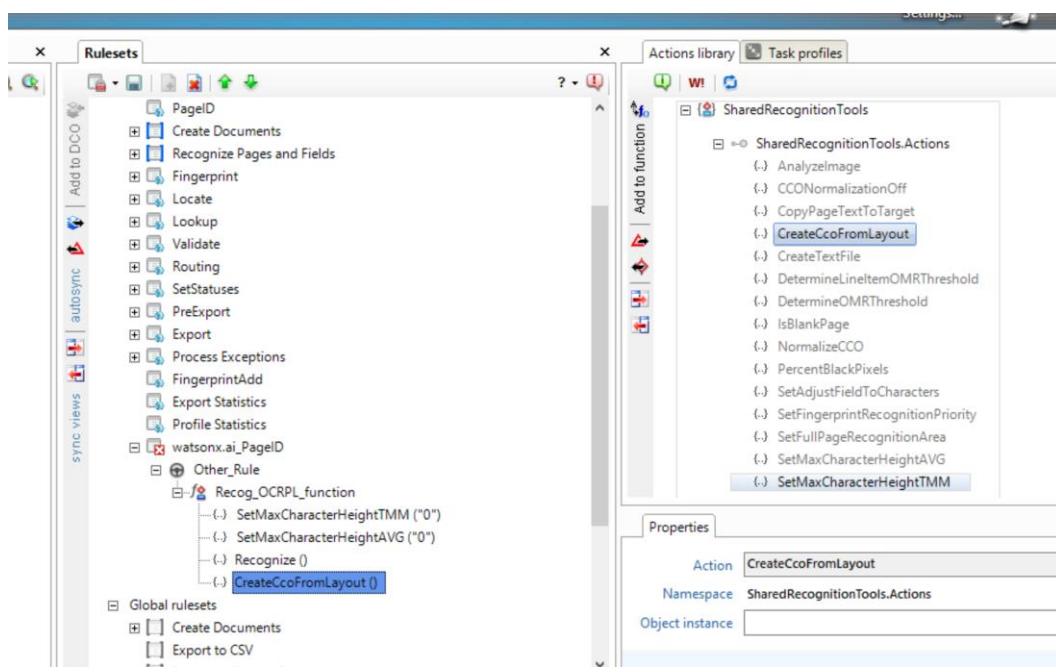
Accuracy in field extraction

Processing speed

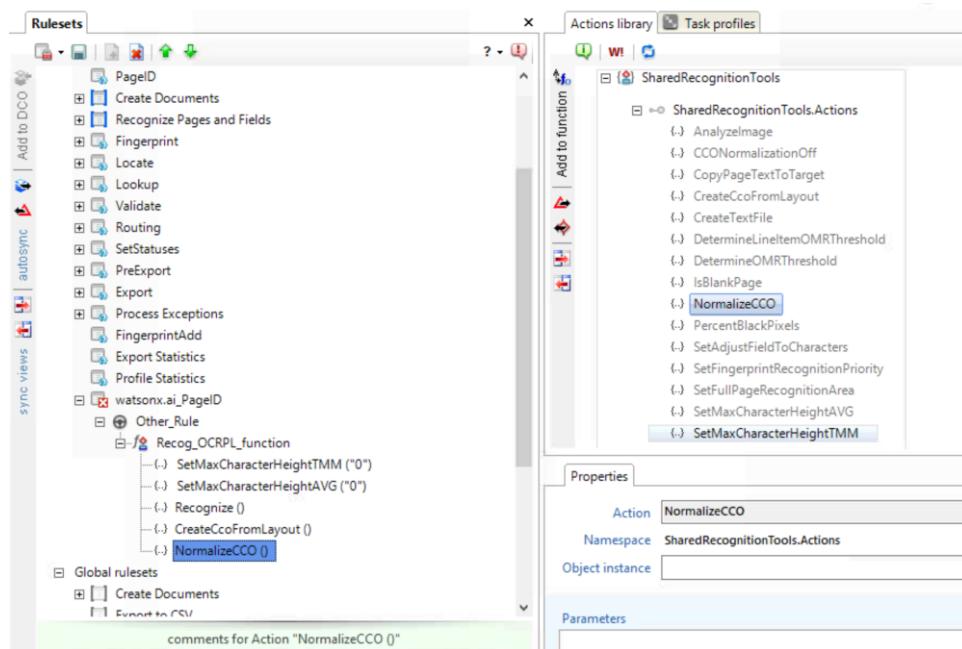
Choosing the right OCR engine is essential for optimizing recognition quality and overall application performance.



\_11. Add the '**CreateCcofromLayout**' method from **sharedRecognitionTools.Actions** under the **SharedRecognitionTools library** to the **Recog\_OCRPL function**.



\_12. Add the ‘NormalizeCCO’ method from **sharedRecognitionTools.Actions** under the **SharedRecognitionTools library** to the **Recog\_OCRPL** function.



Note: Recognize()

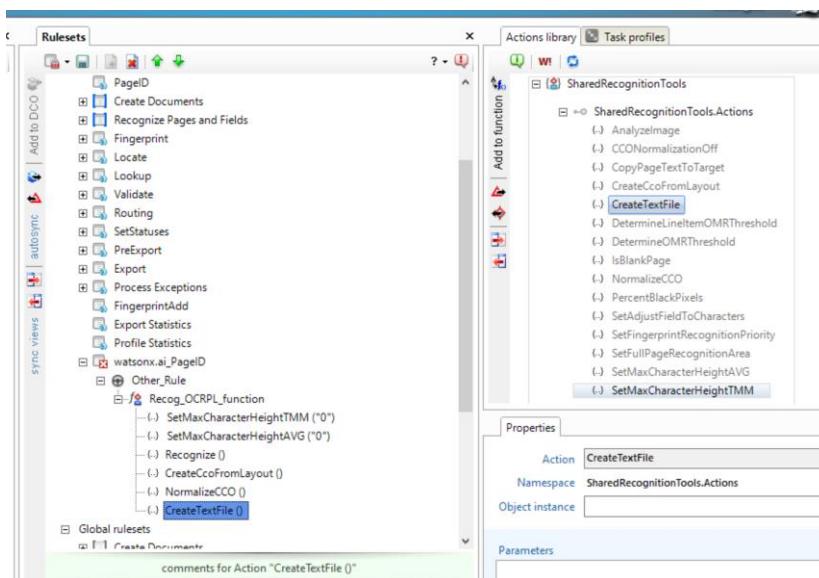
**CreateCcoFromLayout()**

NormalizeCCO()

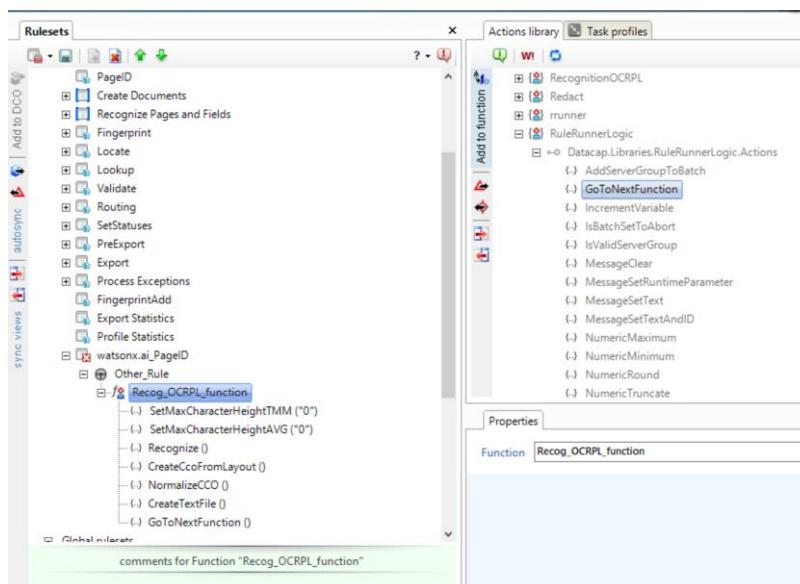
The layout file is created by the Recognize method. Then CreateCcoFromLayout creates the CCO based on the recognition results. NormalizeCCO then normalizes the CCO to get it ready for use by other Datacap clients and actions.

In the context of the IBM Datacap software platform, **CCO** stands for "Content Classification Object" and refers to a file that contains the recognized text and its coordinates on a document. This file is an integral part of the data extraction and validation process.

\_13. Add the '**CreateTextFile**' method from **sharedRecognitionTools.Actions** under the **SharedRecognitionTools library** to the **Recog\_OCRPL function**.



\_14. Add the '**GoToNextFunction**' action from **Datacap.Libraries.RuleRunnerLogic.Actions** under the **RuleRunnerLogic library** to the **Recog\_OCRPL function**.

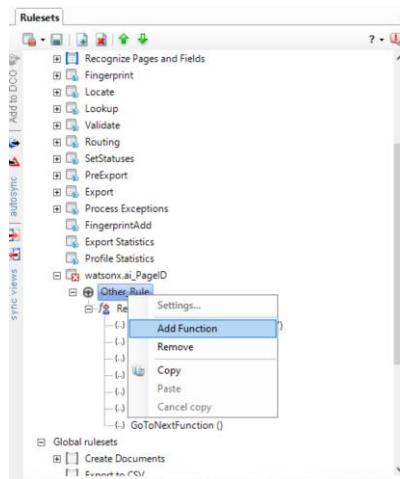


\_15. Click on **Save changes** in Ruleset panel.

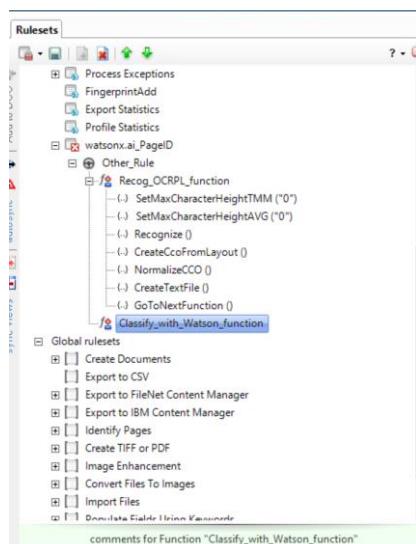
Ensure that all function methods are added in the exact order shown in the screenshot. If your function order differs, you can rearrange them using the green up and down arrows located above the list.

**Note:** To avoid losing any changes made in your ruleset, make sure to click the Save icon regularly in Datacap Studio.

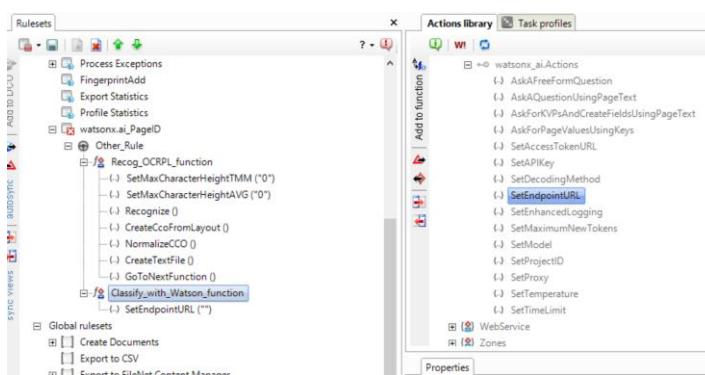
\_16. Add a new function to the **Other Rule** by right-clicking on **Other Rule**.



\_17. Name the new function '**Classify\_with\_Watson\_function**'.



\_18. Add the '**SetEndpointURL**' method from the **Watsonx\_ai.Actions** action under the **watsonx\_ai** library to the **Classify\_with\_Watsonx\_function**.

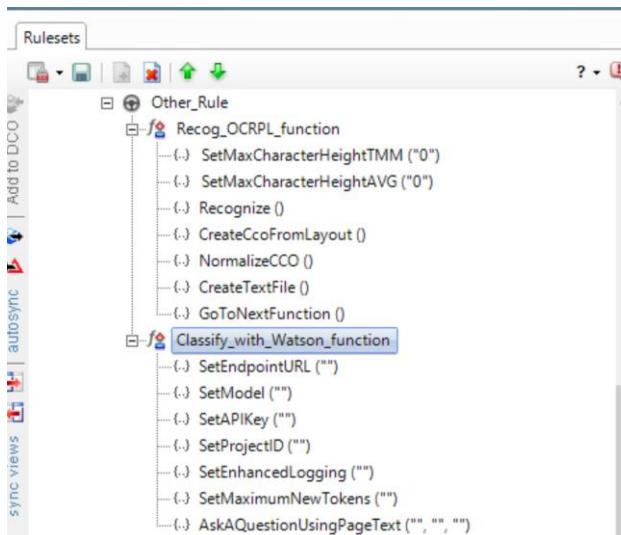


\_19. Add the following methods to the **Classify\_with\_Watson\_function** from the **Watsonx\_ai.Actions** action under the **watsonx\_ai** library following steps mentioned earlier:

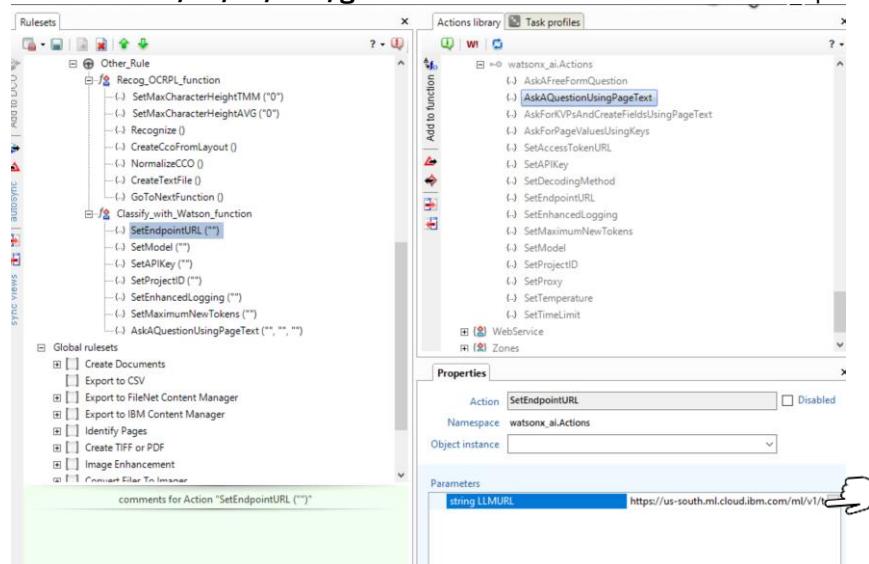
- **SetModel**
- **SetAPIKey**
- **SetProjectID**
- **SetEnhancedLogging**
- **SetMaximumNewTokens**
- **AskAQuestionUsingPageText**

To proceed, you'll need to gather the following details from Watsonx.ai Studio: **Endpoint URL**, **Model Name**, **API Key**, and **Project ID**. You can obtain these by either accessing the environment directly or by following the lab setup instructions starting from section 2.1.1, step 9.

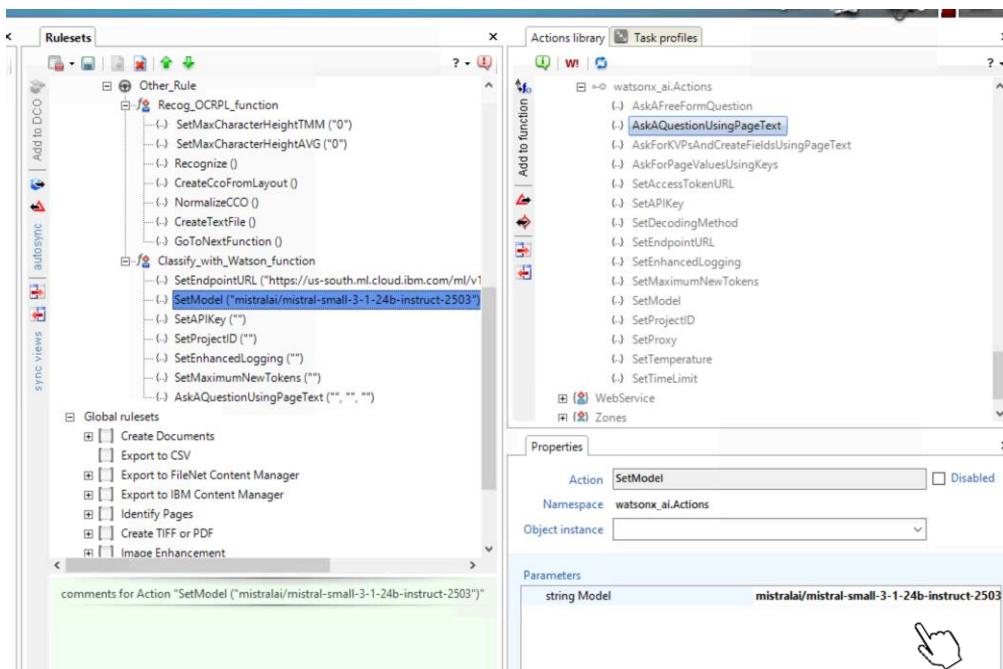
During the TechJam event, these parameters will also be provided. However, please keep in mind that the number of available tokens is limited, so plan accordingly.



\_20. Add the Endpoint URL parameter's value to '<https://us-south.ml.cloud.ibm.com/ml/v1/text/generation?version=2023-05-29>' as a parameter, as shown below.



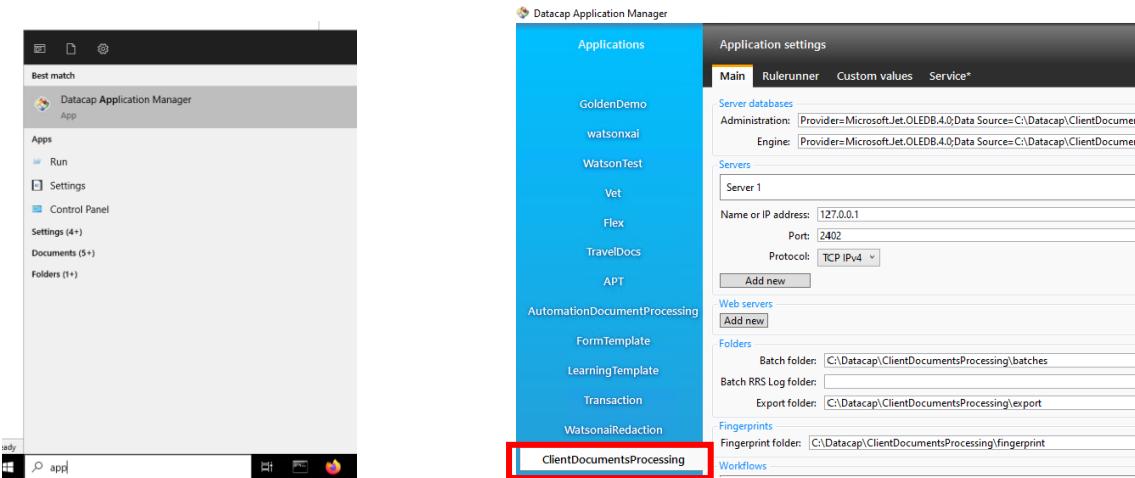
\_21. Set the **model** value to ‘mistralai/mistral-small-3-1-24b-instruct-2503’



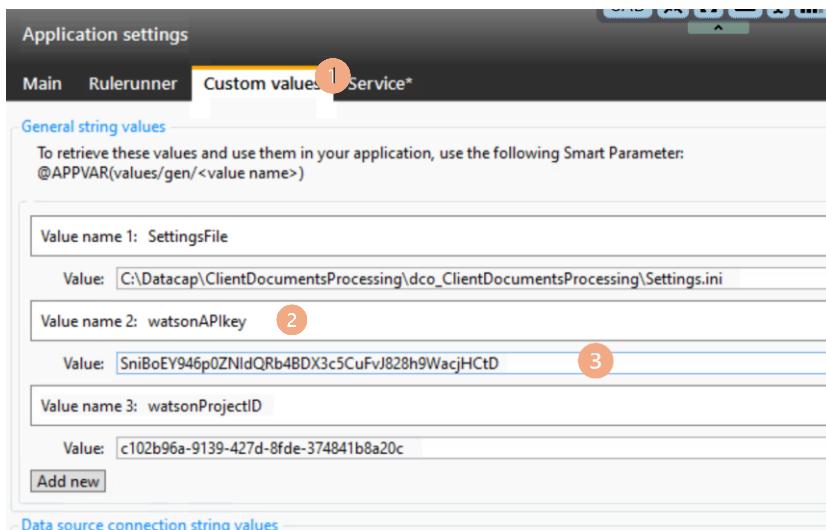
You can configure and use different foundation models available on the Watsonx.ai platform, depending on the region where your project is deployed

The screenshot shows the Datacap Prompt Lab interface. At the top, it says 'Projects / Datacap Watsonx.ai Demo / Prompt Lab'. Below that, there are tabs for 'Chat', 'Structured', and 'Freeform', with 'Freeform' selected. A hint message says 'Hint: This model works better when you provide at least 1 example.' The main area has a text input placeholder 'Enter your prompt text.' and a note: 'Remember: This is not a chat interface. Provide instructions and examples to show the model what to do.' Below that, it says 'When you prompt a text-generating model, the model responds by appending text to your prompt text or continuing your prompt text.' A note also says 'Try the sample prompts for a variety of use cases.' To the right, there's a 'View code' section with a 'Curl' tab selected, showing a complex JSON command for generating text using the specified model. Buttons at the bottom include 'Clear output', 'Generate', and a large '→' button.

\_22. Open the Datacap application manager and click on the **ClientDocumentsProcessing** under **Applications** from the left panel.



\_23. In ClientDocumentsProcessing , navigate to custom values tab, click on add a new variable with **Name: watsonAPIkey** and **value** (*your actual Watsonx.ai APIKey*) under general string values.



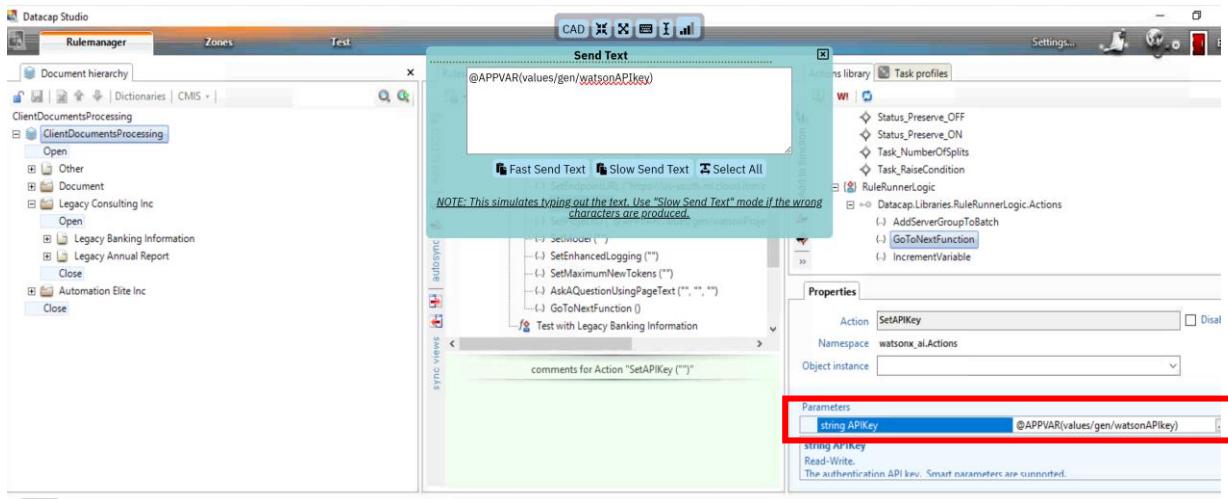
\_24. Add another variable under general string values **Name :watsonProjectID** and **Value:** (*your actual Watsonx.ai Project ID*). Once you add both parameters, save the changes and close it.

\_25. Return to Datacap Studio to set API Key and ProjectID parameters to the variable values.

\_26. Set API Key parameter as a Datacap Application Variable – ‘@APPVAR(values/gen/watsonAPIkey)’ by setting this value in parameter stringAPIKey.

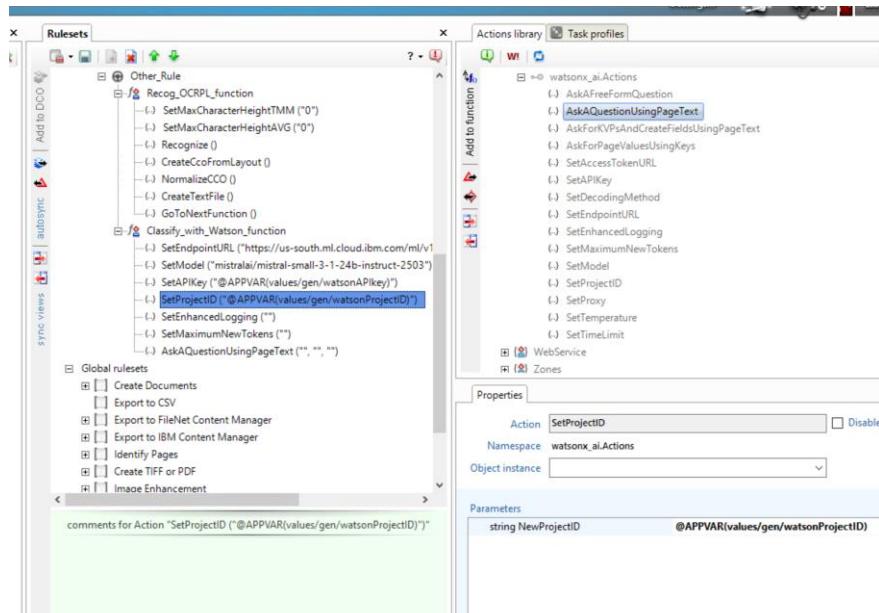
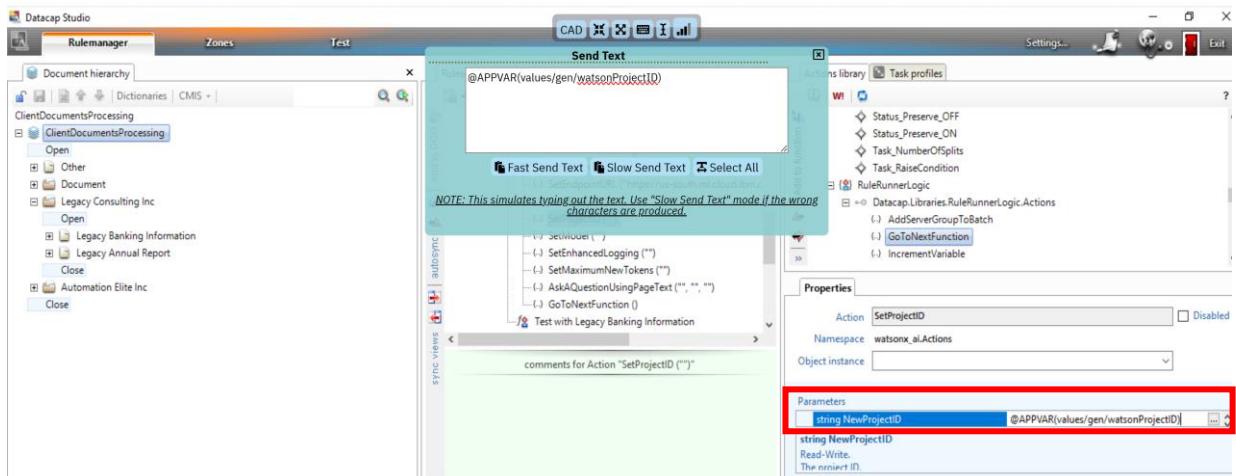
**Note:** To enable integration with Watsonx.ai, you must **add the API Key** associated with your project. This key is used to authenticate and securely access Watsonx.ai services.

Make sure the API Key corresponds to the **correct region and project ID** you are working with.



\_27. Set ProjectID parameter as a Datacap Application Variable '@APPVAR(values/gen/watsonProjectID)'.

**Note:** To integrate your Datacap application with Watsonx.ai, you also need to store the **Project ID** as a variable in?



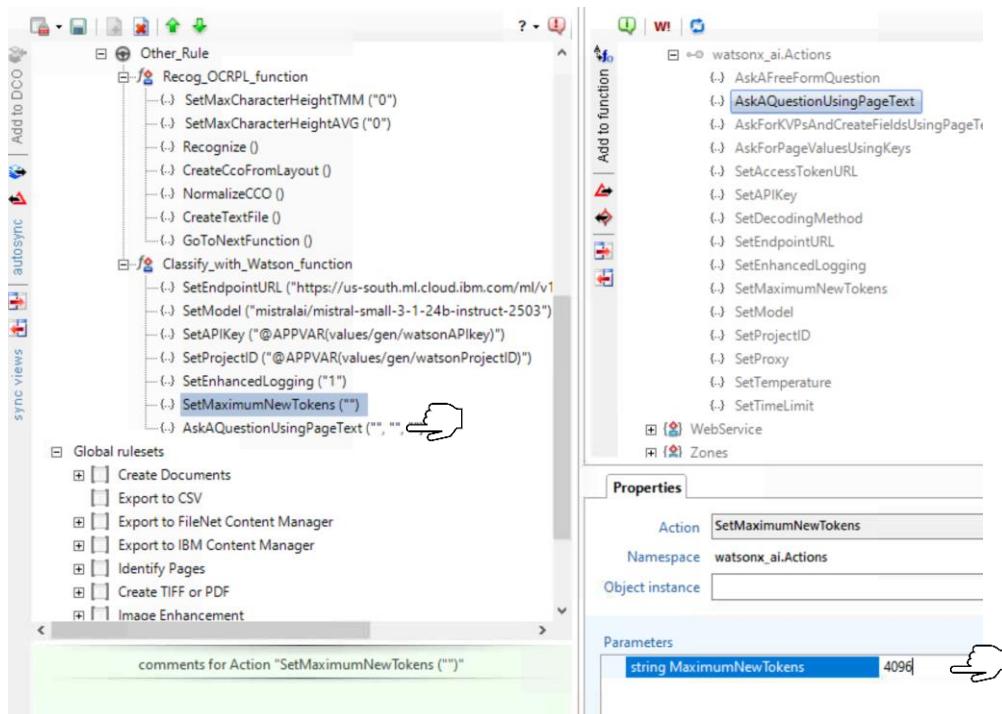
## \_28. Set Enhanced Logging Value to 1.

This setting is helpful to view the prompts sent to Watsonx.ai and the corresponding responses received. Detailed logs will be written in the following files for better understanding.

PageID\_rrs.log

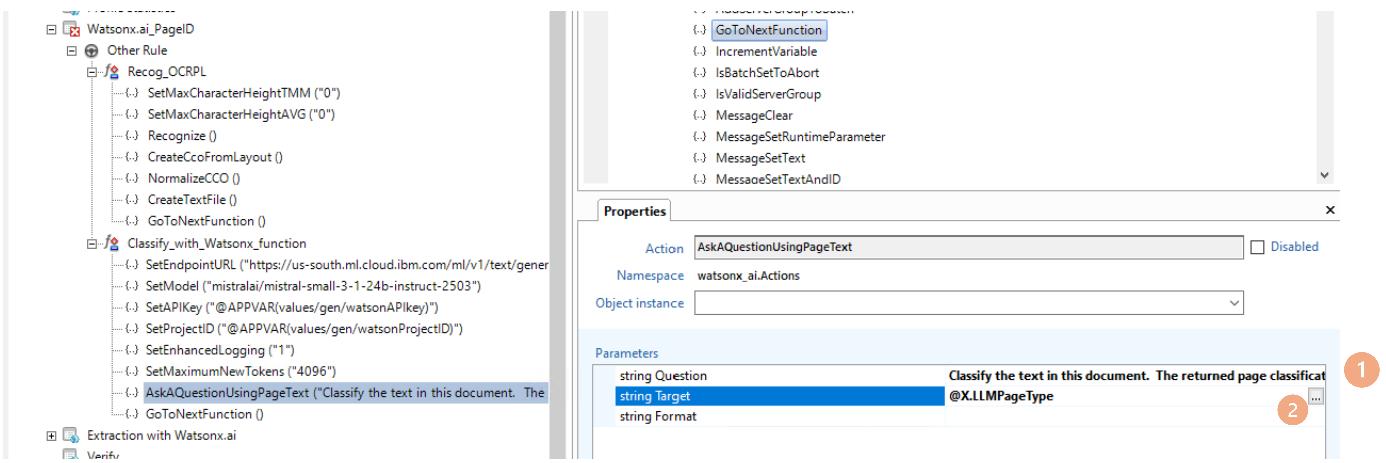
Profiler\_rrs.log

## \_29. Set SetMaximumNewTokens value to 4096 which Controls the number of returned tokens from the LLM.



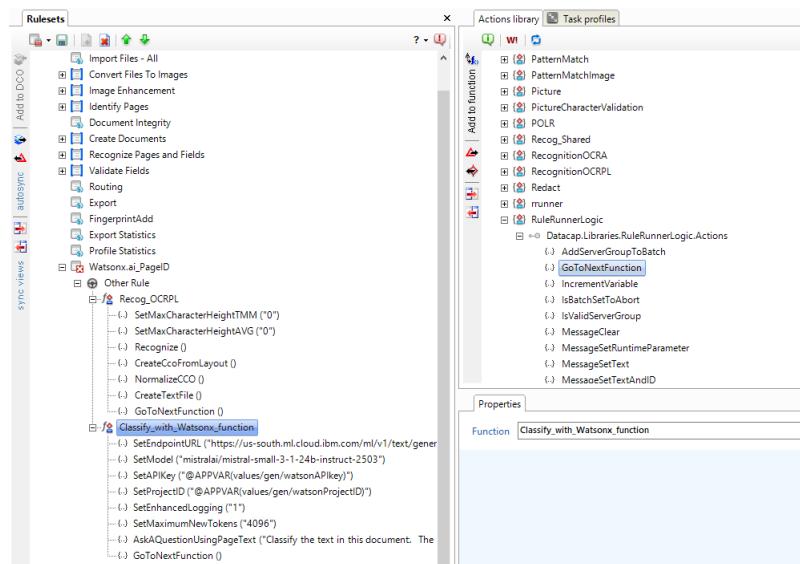
## \_30. Add both string parameter values in AskAQuestionUsingPageText method

- 1) string **Question** = 'Classify the text in this document. The returned page classification must be exactly one of the following classification types: "Legacy Banking Information", "Legacy Annual Report", "Automation Banking Information", "Utility Bill", "Unknown". Only respond with the exact answer. Do not return explanation text.'
- 2) string **Target**= '@X.LLMPageType'

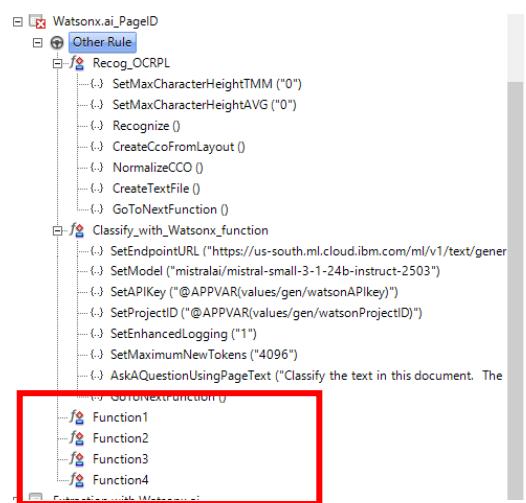
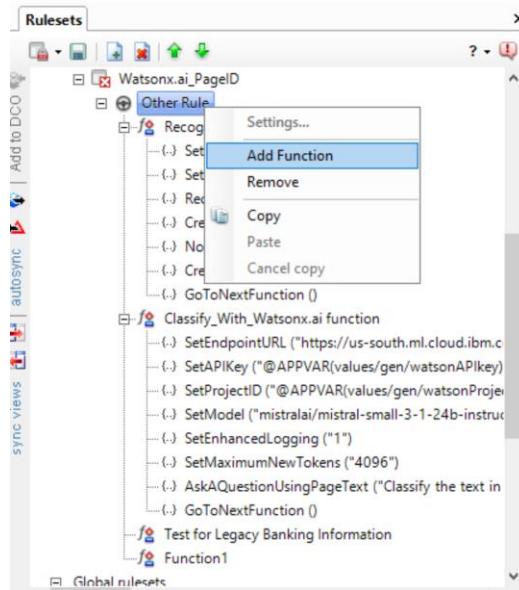


\_31. You can copy the **GoToNextFunction()** method from the **Recog\_OCRPL** function by right-clicking on the method and pasting it into the **Classify\_with\_Watsonx\_function**.

Alternatively, you can add this method from the **Datacap.Libraries.RuleRunnerLogic.Actions** under the **RuleRunnerLogic** library.

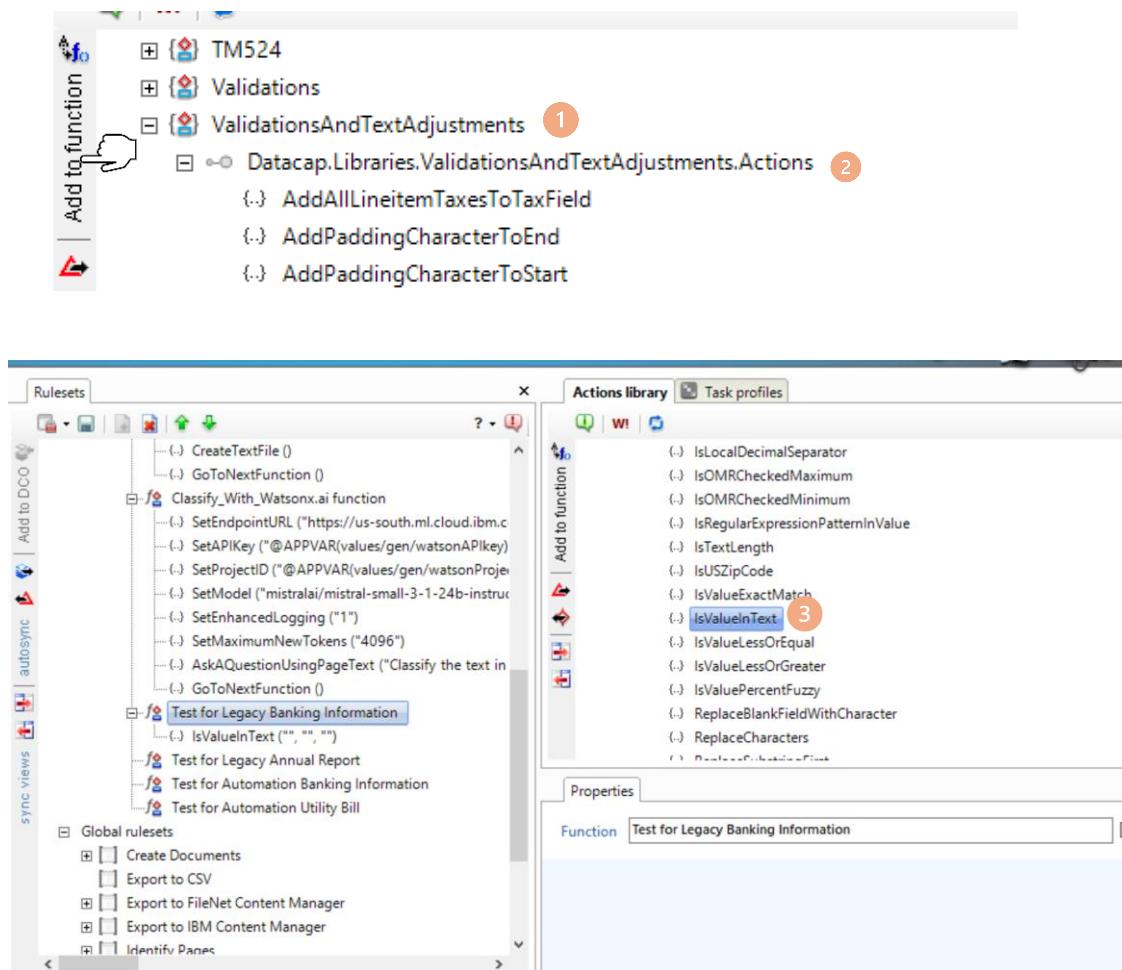


\_32. Add four new functions by right-clicking on the **Other Rule** four times. These functions will be used to set the Page Type dynamically based on the **LLMPageType** response returned from the Page Classification process.



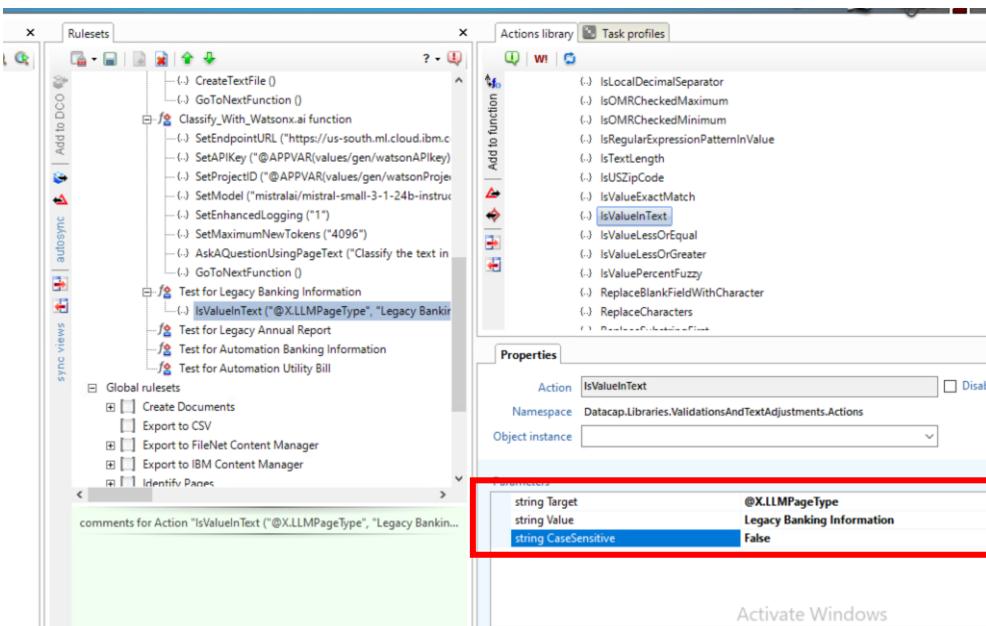
\_33. Rename the four newly added functions as shown below.

\_34. Select the **Test for Legacy Banking Information** function, then add the **IsValueInText** method by selecting it from the action library **ValidationsandTextAdjustments** and Datacap libraries and clicking **Add to Function**.

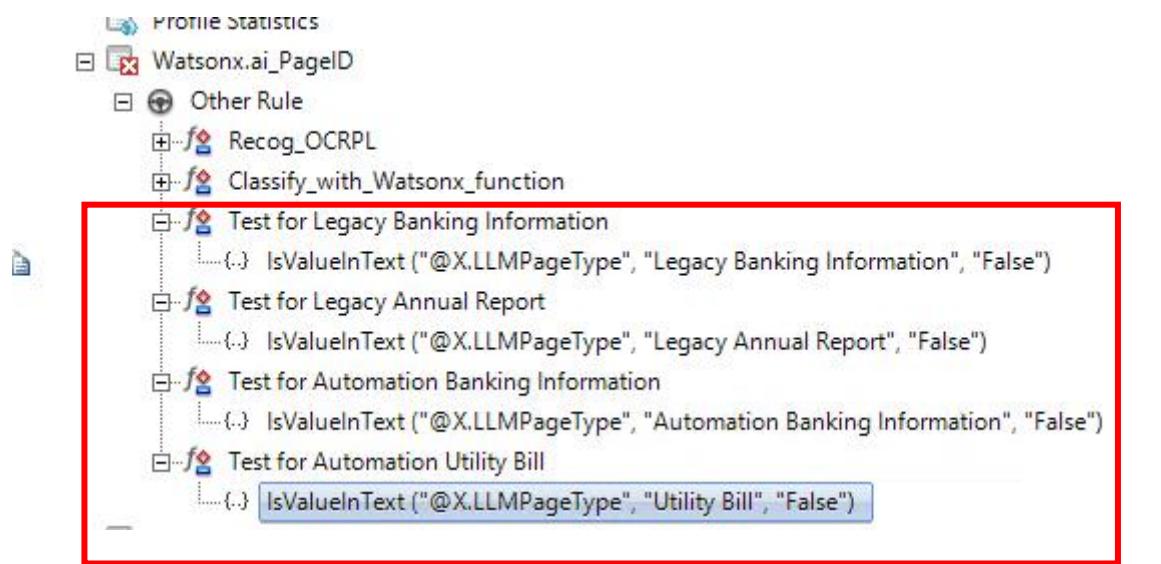


\_35. Set the following parameter values for the **IsValueInText** method to check if the return response matches the expected value:

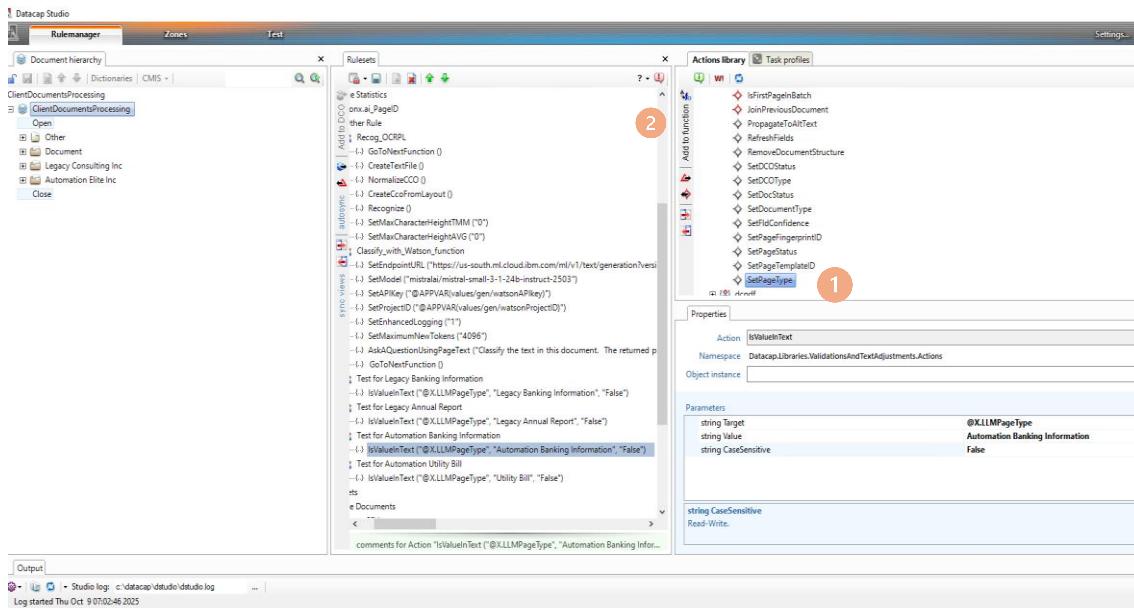
- **Target string value:** @X.LLMPageType
- **String value:** Legacy Banking Information
- **CaseSensitive:** False



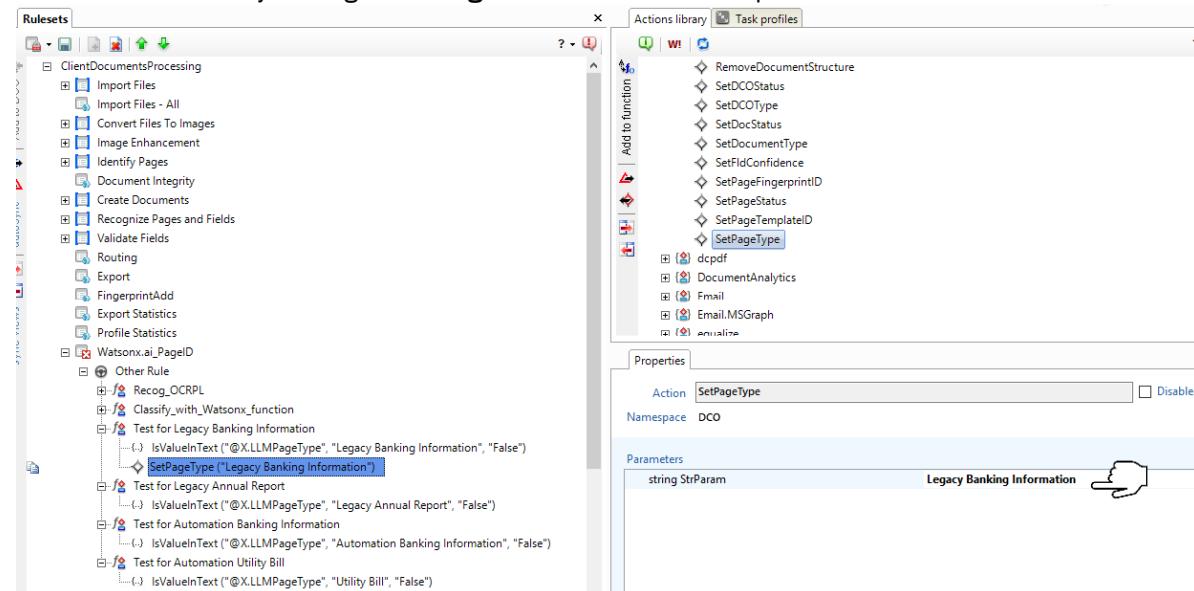
\_36. Add the **IsValueInText** method along with the corresponding parameters to the remaining three functions, as shown below.



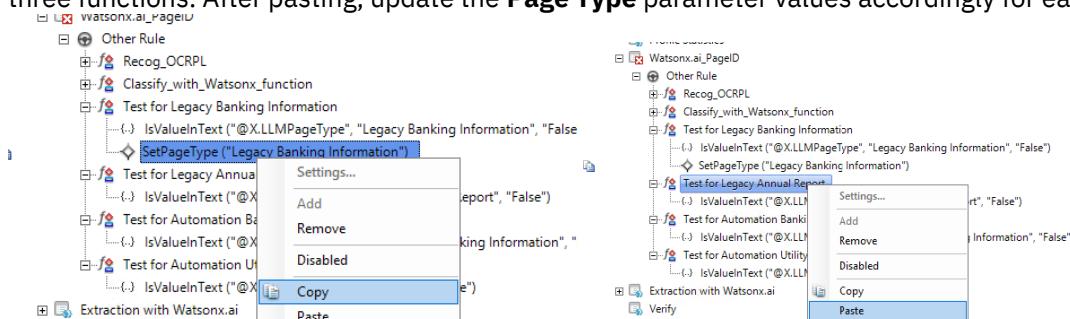
\_37. After adding the **IsValueInText** method and its parameters to all four functions, select the **Test for Legacy Banking Information** function and add the **SetPageType** method to it from the **DCO** action library.



**\_38. Set the Page Type parameter value to ‘Legacy Banking Information’ for the Test for Legacy Banking Information function by setting the string StrParam value in Properties.**

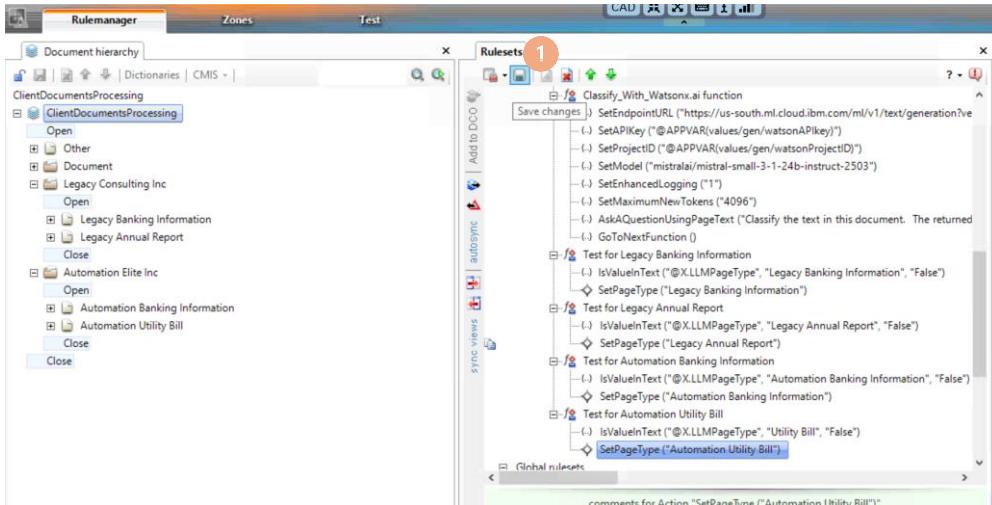


**\_39. You can copy the SetPageType method by right-clicking on it (as shown below), then paste it into the remaining three functions. After pasting, update the Page Type parameter values accordingly for each function.**

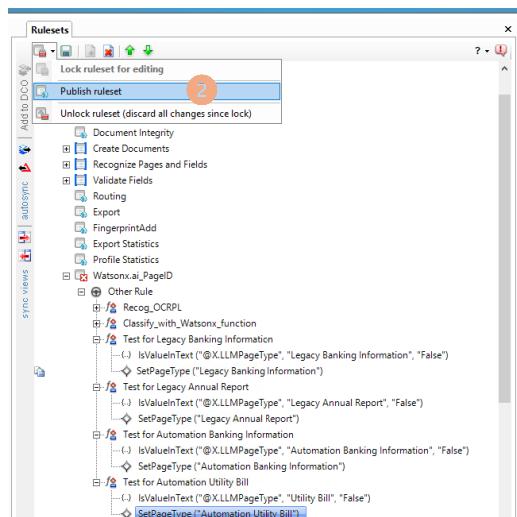


The order of methods in a function matters as they will be run sequentially. So, once you add the methods inside function, check from the screenshots to see if the order appears as is displayed. This will ensure that the steps are executed sequentially and no issues are faced.

- \_40. Click the **Save** icon and then **Publish** the ruleset, as shown in the screen below.



You have successfully created the ruleset for **recognition** and **page classification** using **Datacap Studio**.



### 3.2.2.2 Map Watsonx.ai\_PageID Ruleset to Task Profile Workflow

1. Go to **Task Profiles** in right panel and open PageID. Next, remove all other rulesets from the **PageID** task under **Task Profiles**, except for the **CreateDocuments** and **Document Integrity** rulesets.

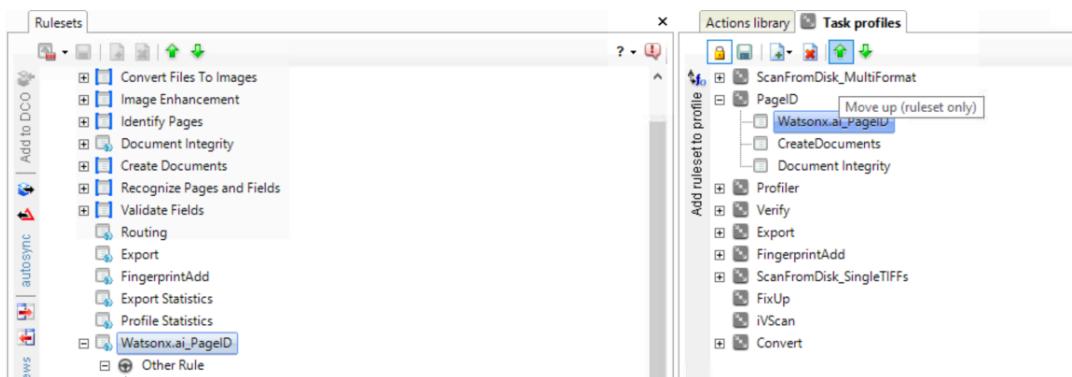
Please note that the previous step instructed you to remove the default rulesets from the **PageID** task profiler to streamline execution and eliminate unnecessary steps. Ensure that only the following three rulesets remain, in this exact order:

**Watsonx.ai\_PageID**

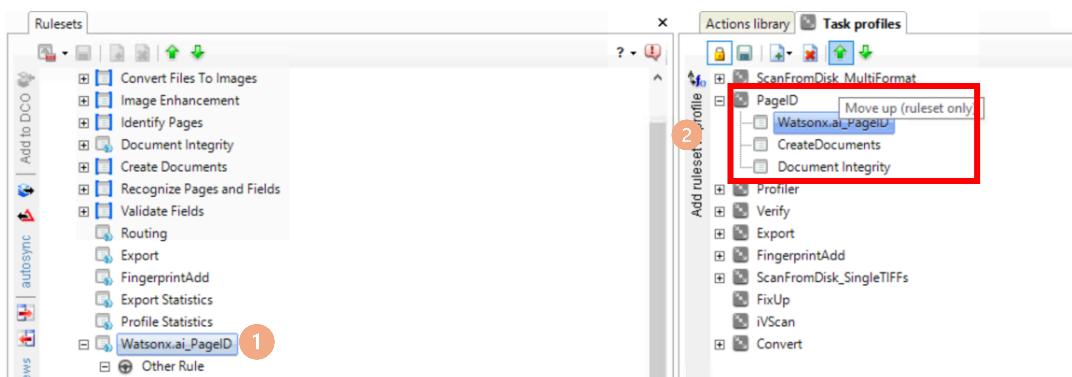
**CreateDocuments**

**Document Integrity**

If you'd like to understand the purpose of [CreateDocuments](#) and [Document Integrity](#), refer to the official IBM documentation for detailed explanations.



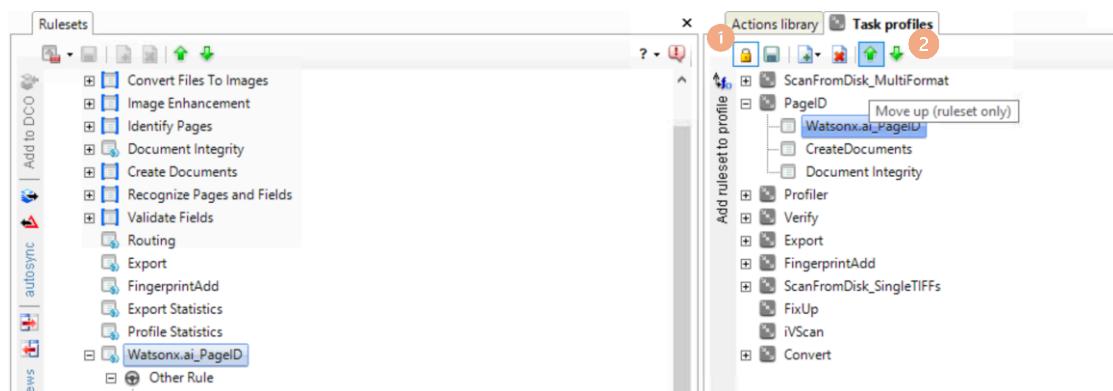
- Select Watsonx.ai\_PageID from the Rulesets panel and click on **Add ruleset to profile** button to Add the Watsonx.ai\_PageID ruleset to the **PageID Task Profile** workflow under the **PageID** section.



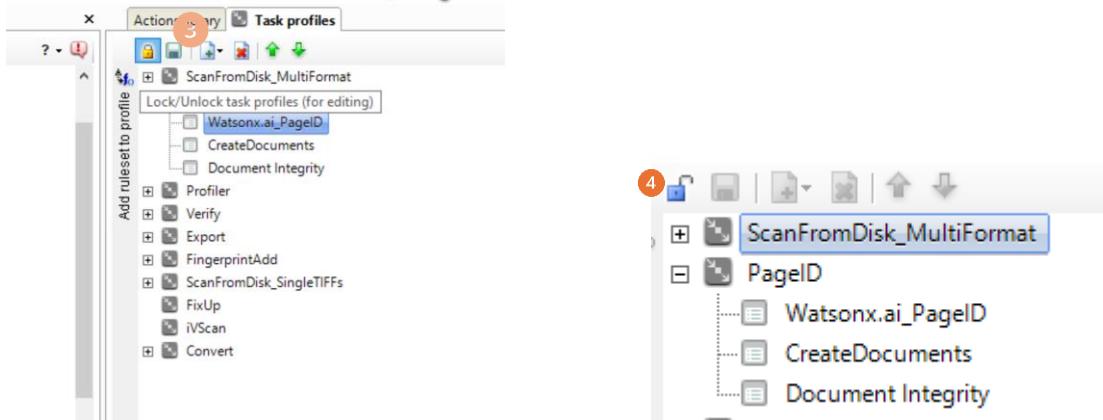
Make sure to add the Watsonx.ai\_PageID ruleset *before* the CreateDocuments task, as shown below.

Before editing Task Profiles, you must lock them by clicking the lock icon on the Task Profiles tab.  
To remove a ruleset, click the "X" icon next to the ruleset name.

To add a ruleset, click the "Add Ruleset to Profile" button and select the desired ruleset from the list.



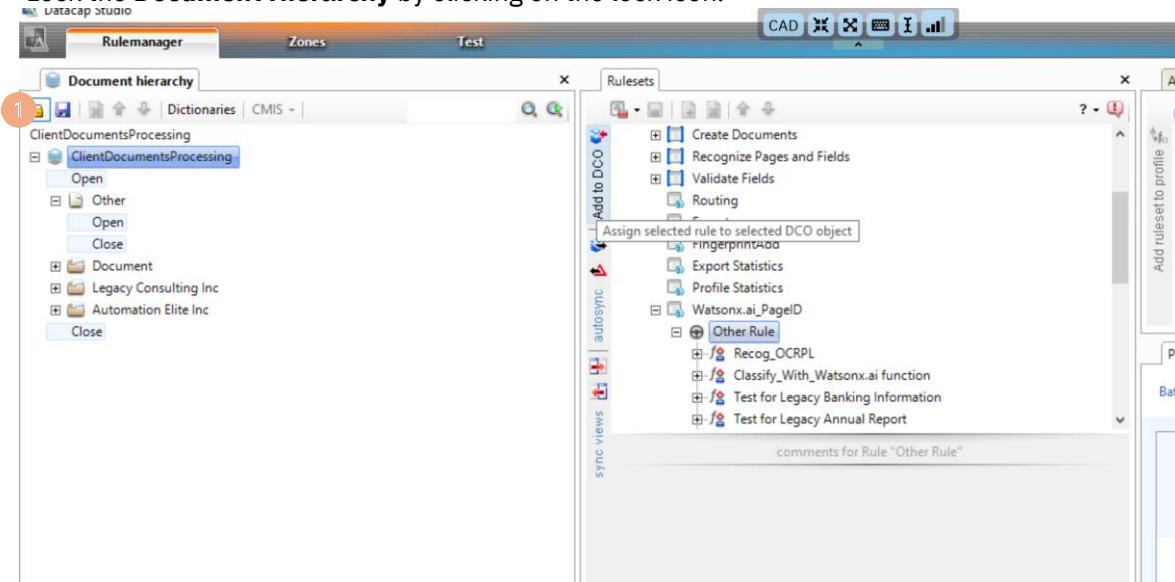
To make changes in the task profiles, lock the Task profiles, move the task profiles up or down as per requirement and lock the Task Profiles again.



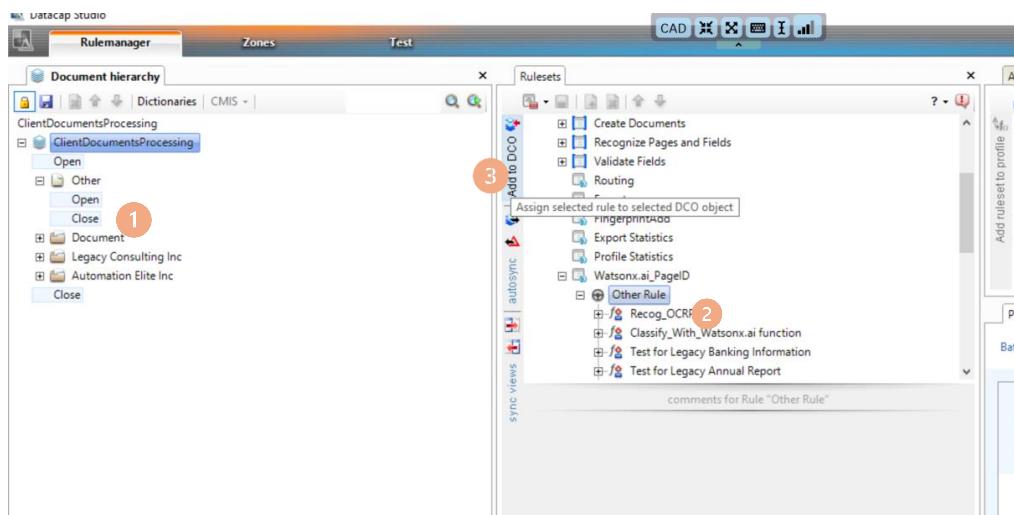
### 3.2.2.3 Map the Watsonx.ai\_PageID (Other Rule) to the Document Hierarchy (DCO)

Now, add the rule to the **DCO (Document Hierarchy)** to specify where the ruleset should be executed during processing.

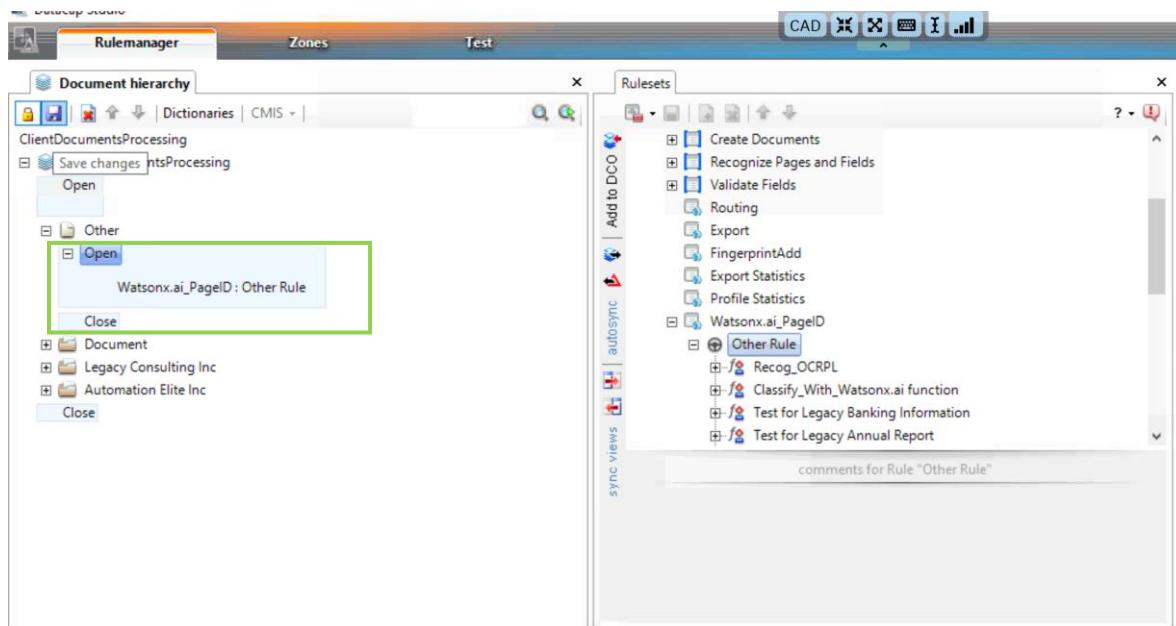
1. Lock the **Document Hierarchy** by clicking on the lock icon.



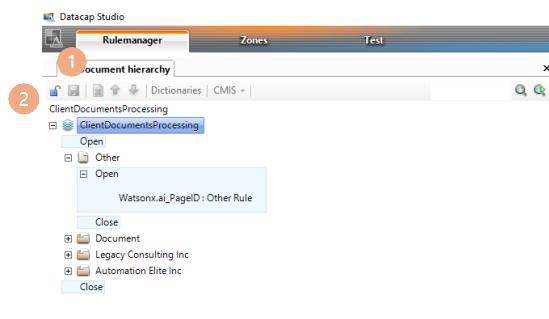
2. Expand the **Other** in document hierarchy pane, select Open, we want to associate the rule to Other document hierarchy. Click on **Other Rule** from the ruleset under Watsonx.ai\_PageID and click on '**Add to DCO button**' to apply this rule to Document Class.



After the rule added to the **DCO** object, you should see it reflected as shown in the screenshot below.



3. Click **Save** icon, then click the **lock** icon to **unlock** the **DCO** as shown below.



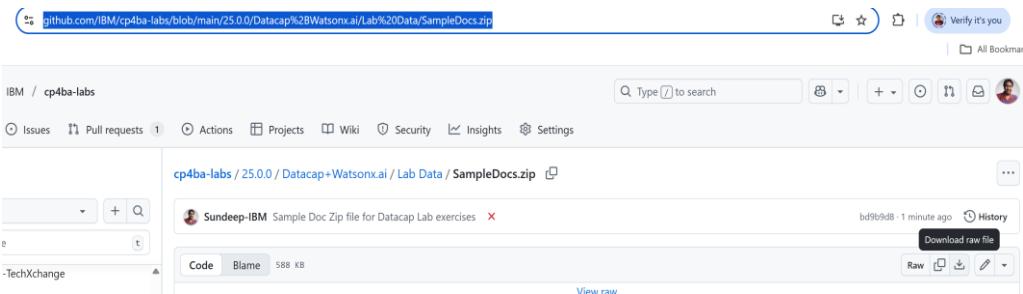
You have successfully enhanced the Datacap application to recognize documents using the OCRPL (Planet.AI) engine and to perform Document classification with Watsonx.ai, enabling accurate document classification prior to extraction.

IBM Datacap, when integrated with watsonx.ai, transcends the limitations of traditional OCR systems by introducing **dynamic layout interpretation** and **semantic understanding**. Traditional OCR relies on fixed templates and basic character recognition, often struggling with unstructured or variable document formats.

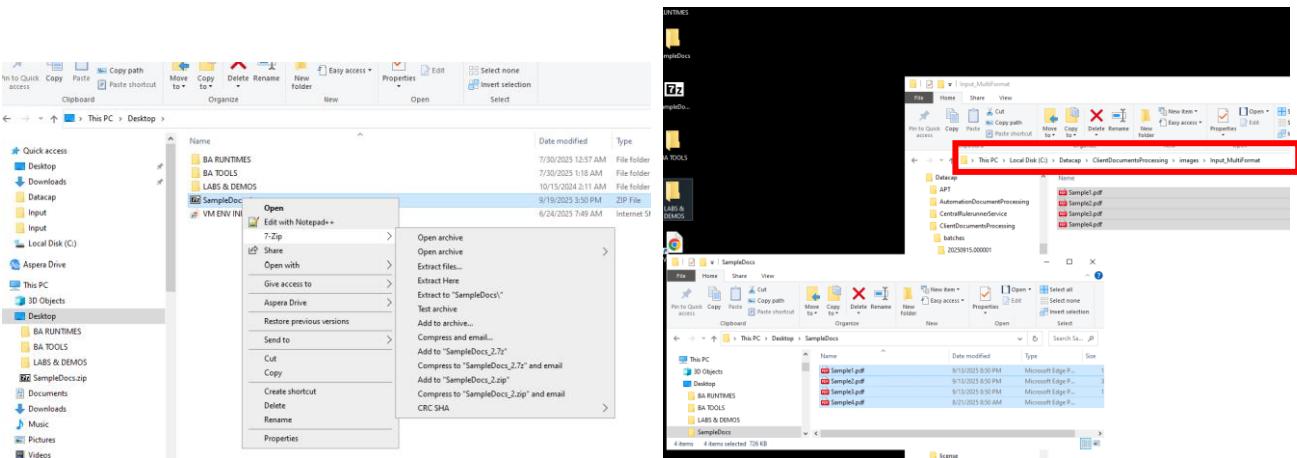
### 3.2.2.4 Verify/Test the Document classification within DataCap studio with Client Documents

In this section you are validating the Watsonx.ai\_PageID ruleset in DataCap with Client Documents.

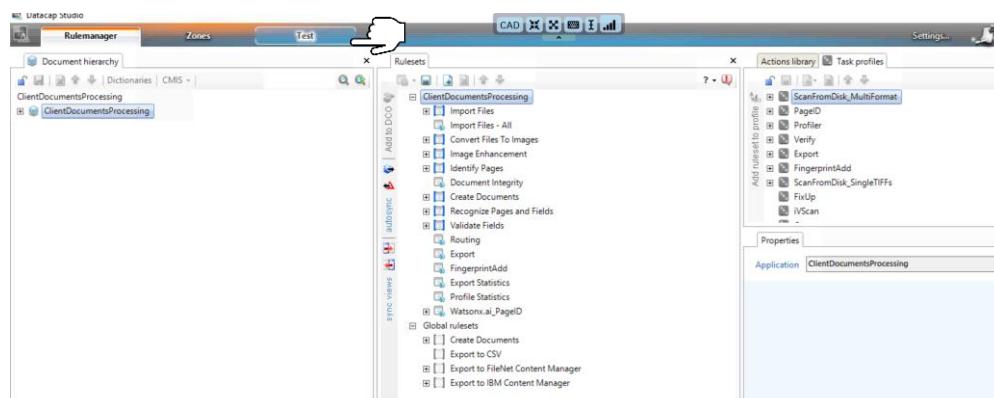
- \_1. Download the sample documents into the VM from the GitHub link <https://github.com/IBM/cp4ba-labs/blob/main/25.0.0/Datacap%2BWatsonx.ai/Lab%20Data/SampleDocs.zip> as shown below.



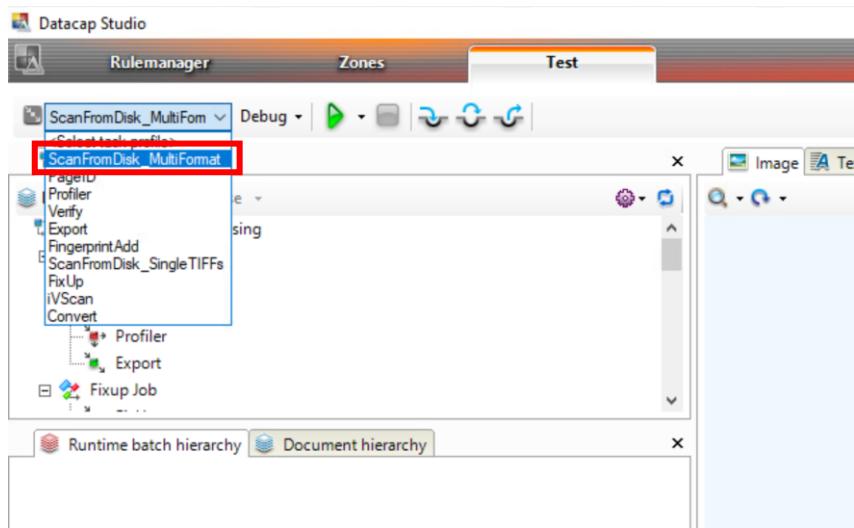
- \_2. Unzip the SampleDocs folder and copy the sample documents in to Input folder "**C:\Datacap\ClientDocumentsProcessing\images\Input\_MultiFormat**".



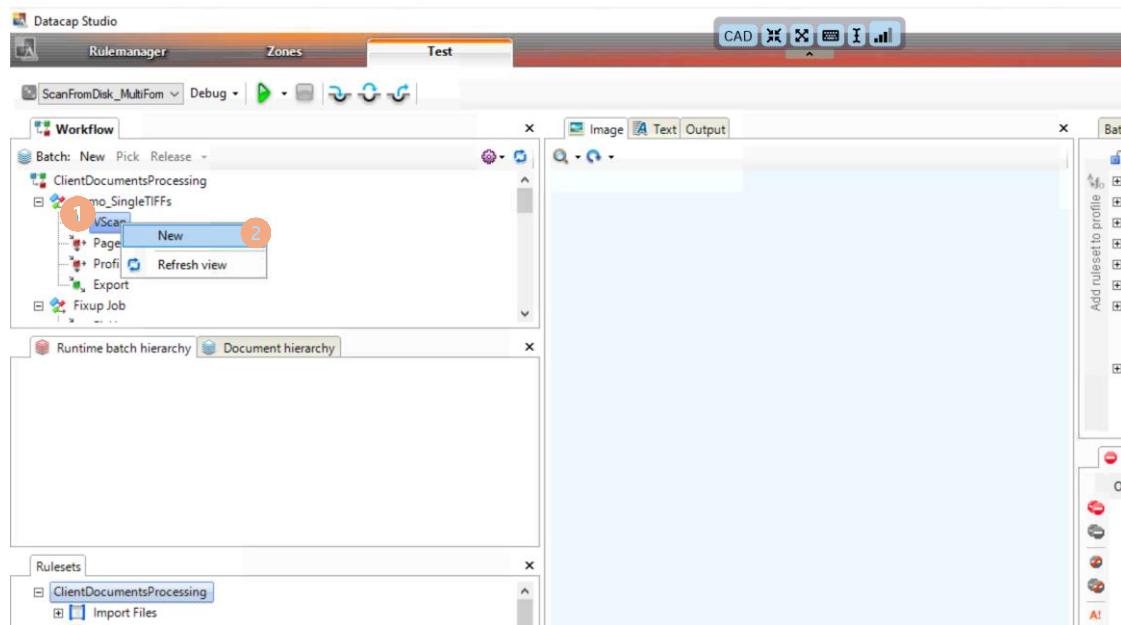
- \_3. Click the **Test** tab from the DataCap studio



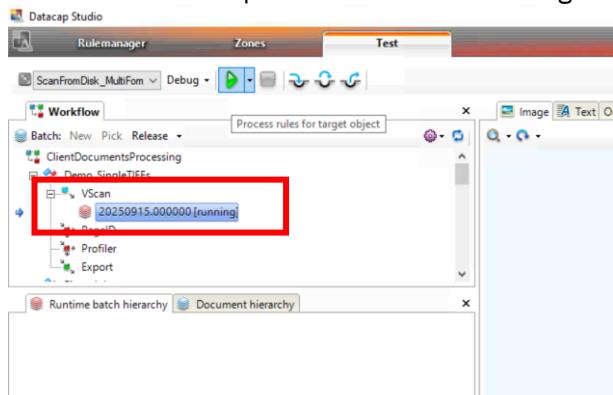
\_4. Select **ScanFromDisk\_MultiFormat** from the **dropdown** menu as shown below.

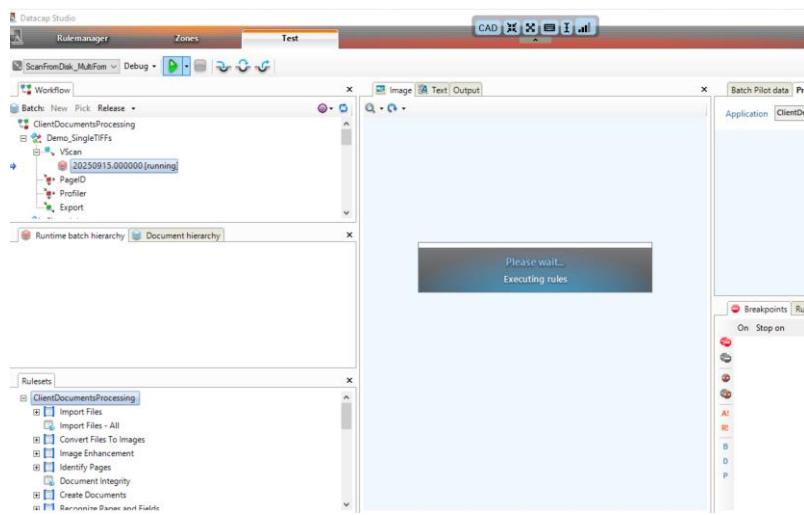


\_5. Right click on the Vscan under **Demo\_singleTiffs** and click **New** and click the green **Play** button to start the process.



\_6. Now the **vScan** task profiler rules will scan the given input sample docs.



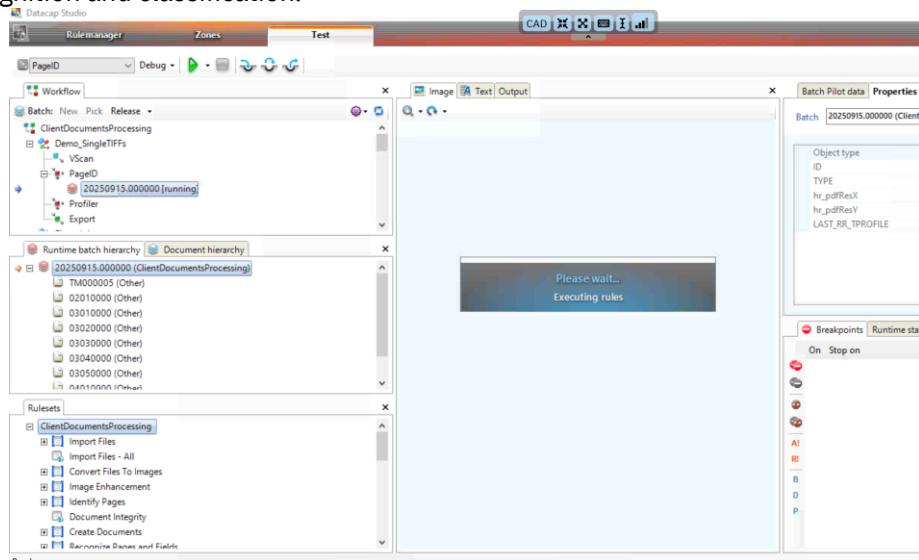


In the Vscan task-All the sample documents get converted from pdf into tiff images and moved to Other Page. Open File Explorer, navigate to C:\Datacap\ClientDocumentsProcessing\Batches\, and select the current batch folder to observe behind-the-scenes activity during the Vscan task.

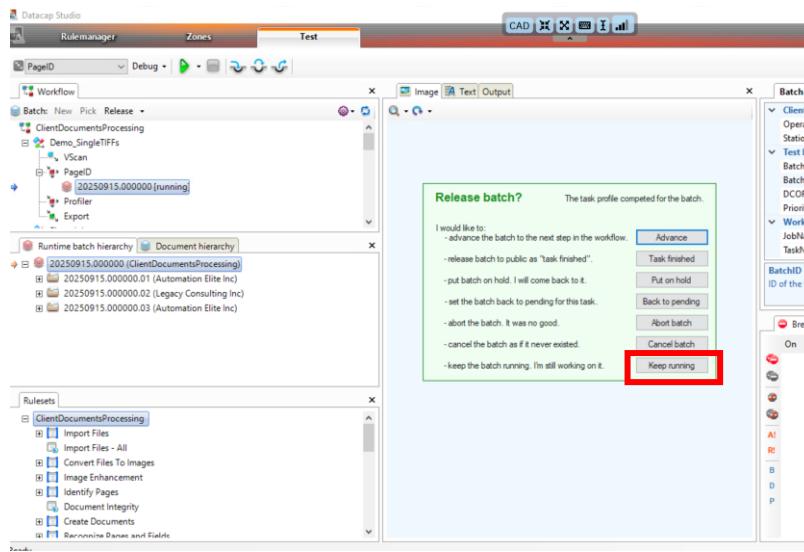
If you encounter an error during task execution, you can check the corresponding RRS logs—such as **Vscan\_rrs.log**—in the current batch folder located in File Explorer.

\_7. Click on the “**Advance**” button from the popup window and click the green **Play** button to start the process.  
<Screenshot to be added here>

\_8. Now the batch has moved to **PageID task profiler** where you added the watsonx.ai\_PageID ruleset for recognition and classification.



\_9. When the popup window appears, click the “Keep Running” button. This allows you to verify the results before moving forward. It's a recommended practice to check outcomes at each stage to ensure smooth execution and accurate results.



You can see the documents are successfully classified based on the Watsonx.ai LLM model response. In case, documents are not classified then go to troubleshooting section to troubleshoot your application.

10. Expand each **Document Type** in the **Runtime Batch Hierarchy** and verify how each page has been classified to ensure the Document classification logic has been applied correctly.

The screenshot shows the Datacap Studio interface with the 'Test' tab selected. On the left, the 'Workflow' panel shows a batch named '20250915.000000 [running]'. The 'Runtime batch hierarchy' panel (1) shows a tree structure of document types and their sub-items. The 'Image' tab (2) displays a scanned document page from 'CABLEPORIUM' showing service charges for a digital TV plan. The 'Text' tab shows the detailed breakdown of charges:

Service Period	Description	Amount
00424-98452	Cable TV Plan - Digital HD	\$ 9.11
	Includes: Digital Starter Programming, Interactive Program Guide, Music Choice	\$ 7.38
	<b>Total Base Cable Fee</b>	<b>\$ 16.41</b>
	Additional Services	
	Pay Per View Service	\$ 8.85
	Additional Outlet - Digital Converter	\$ 8.76
	<b>Total Additional Service Charges</b>	<b>\$ 18.43</b>
	Other Charges	
	Broadcast TV Fee	\$ 11.26
	Regional Sports Fee	\$ 10.06
	<b>Total Other Charges</b>	<b>\$21.32</b>
	Taxes & Surcharges	
	State Sales Tax	\$ 3.75
	Franchise Related Fee	\$ 2.85
	<b>Total Taxes &amp; Surcharges</b>	<b>\$6.60</b>
	<b>TOTAL CHARGES FOR THE PERIOD</b>	<b>\$ 62.76</b>

At the bottom of the document page, there is a 'CONTACT US' section with a phone number and website address.

### 3.2.2.5 Document Extraction using Watsonx.ai Large Language Models (LLM)

Create Extraction with Watsonx.ai Ruleset (Rules and Functions) for the ClientDocumentsProcessing App using Datacap Studio

In this section, you will create two rulesets—“**Extraction all fields with Watsonx.ai**” and “**Extraction values by defined fields with Watsonx.ai**”—including the relevant rules and functions, as illustrated in the diagram below.

Note:

- **Extraction All Fields with Watsonx.ai:** This ruleset automatically identify and extract **all possible fields** from a document—without predefined rules or templates.

Significance:

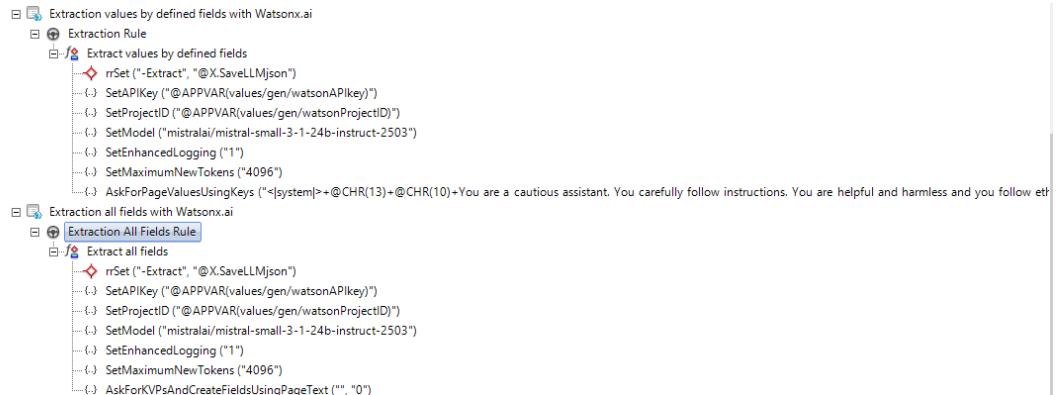
- o **Fully AI-driven:** Relies on foundation models to understand document structure and content.
- o **Flexible:** Ideal for unstructured or semi-structured documents where field positions may vary.
- o **Fast setup:** No need to define fields manually, Watsonx.ai infers them.
- o **Use case:** Best for exploratory tasks, rapid prototyping, or when dealing with diverse document formats.

Note:

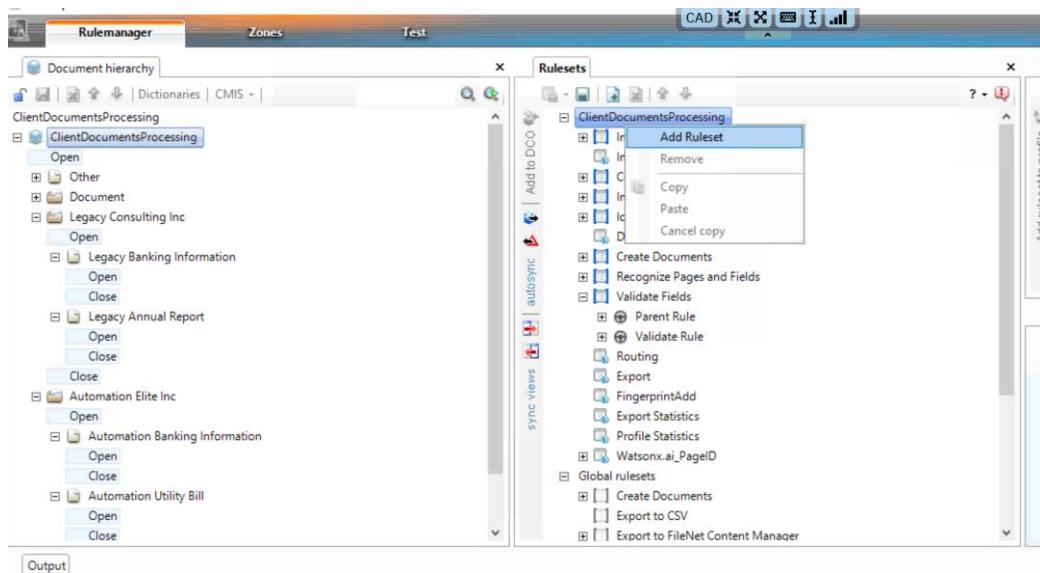
- **Extraction Values by Defined Fields with Watsonx.ai:** This ruleset to extract values **only for specific fields** that have been defined in advance (e.g., Client name, Account number, Address etc).

Significance:

- o **Targeted extraction:** Focuses on retrieving values for known, relevant fields.
- o **Higher accuracy:** AI is guided by field definitions, reducing ambiguity.
- o **Controlled output:** Ensures consistency across batches and aligns with business rules.
- o **Use case:** Ideal for production workflows, compliance-driven processes, or when integrating with downstream systems like Datacap.



\_1. Add a new ruleset by right-clicking on **ClientDocumentsProcessing** and click on **Add Ruleset**.



\_2. Rename Ruleset1 to '*Extraction all fields with Watsonx.ai*' by right clicking and selecting option to rename, rename Rule1 to '*Extraction All Fields Rule*', and rename Function1 to '*Extract all fields*', as shown below



\_3. Copy the methods from **classify\_with\_Watsonx\_function** under the **Watsonx.ai\_PageID** ruleset and paste them into the **Extract all fields** function. Alternatively, you can add these methods directly from **watsonx.ai** action library and set the corresponding parameters.

Make sure the **APIKey** and **ProjectID** methods are set **before** the **Model** method within the function; otherwise, you may encounter errors when calling Watsonx.ai.



\_4. Add the **AskForKVPsAndCreateFieldsUsingPageText** method from watsonx.ai action library to the function and set the **String format** parameter value to **0**.

Note: The function **AskForKVPsAndCreateFieldsUsingPageText** in IBM Datacap is used to extract key-value pairs (KVPs) from the page text and automatically create corresponding fields in the document hierarchy. It leverages AI capabilities—such as Watsonx.ai—to analyze the page content and identify

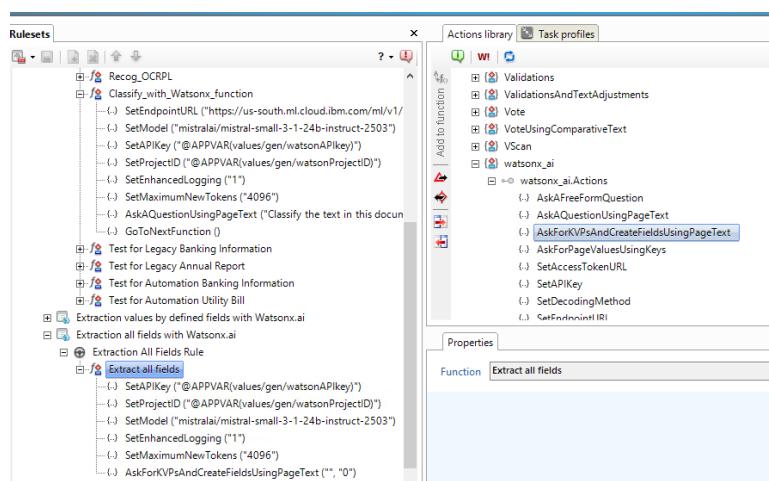
relevant data points without predefined templates. For full technical details and usage examples, refer to IBM's official Datacap API documentation [AskForKVPsAndCreateFieldsUsingPageText - IBM Documentation](#)

To ensure accurate results, you must add all required methods in the correct sequence as shown in the reference screen. Missing or misordered methods may lead to incomplete or incorrect field extraction.

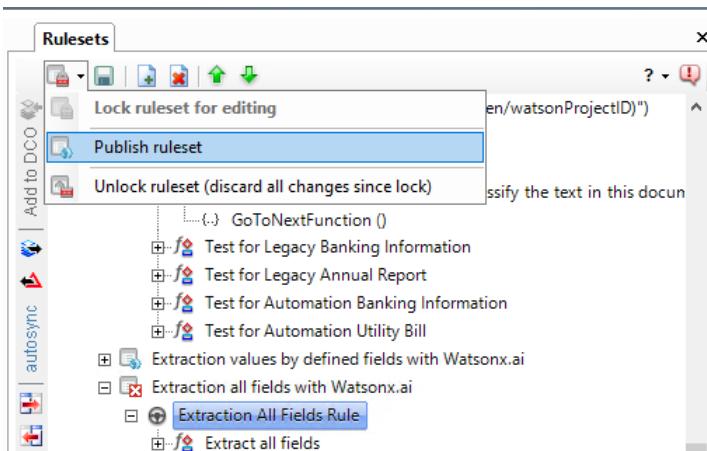
In IBM Datacap, the **String Format** parameter determines how the page content is sent for processing:

- **0** → Sends the page as **plain text** format
- **1** → Sends the page as **HTML** format

This setting affects how functions like AskForKVPsAndCreateFieldsUsingPageText interpret and extract data from the page. Use format **0** for simpler text-based documents, and format **1** when layout or styling (like tables or labels) is important for accurate extraction.

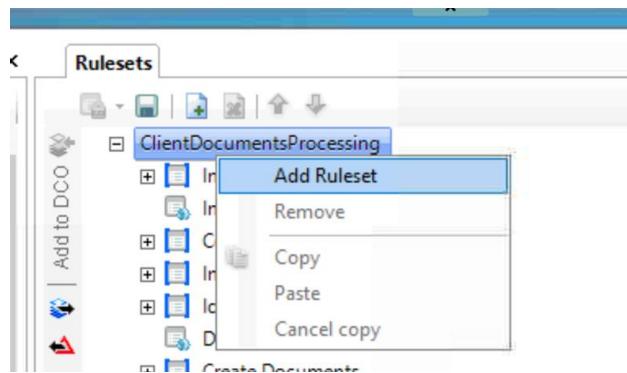


\_5. Save and publish the Extraction all fields with Watsonx.ai ruleset.

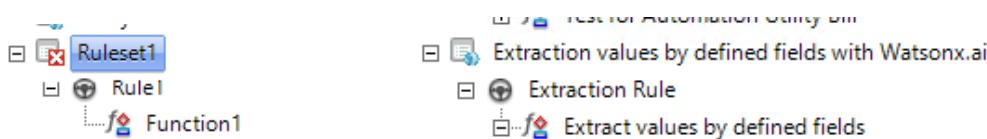


Add a new ruleset to the **ClientDocumentsProcessing** application to extract the defined fields using a Watsonx.ai model, leveraging the **AskForPageValuesUsingKeys** method from the Watsonx.ai Action Library.

\_6. Right click on the **ClientDocumentsProcessing** and click on **Add Ruleset**.



\_7. Rename **Ruleset1**, **Rule1**, and **Function1** to ‘*Extraction values by defined fields with watsonx.ai*’, ‘*Extraction Rule*’, and ‘*Extract values by defined fields*’, respectively, as shown below.



\_8. Copy the methods **SetAPIKey**, **SetProjectID**, **SetModel**, **SetEnhancedLogging**, and **SetMaximumNewTokens** from the earlier ruleset, then add the **rrset** method from the RuleRunner action library, setting **varSource** to “-Extract” and **varTarget** to @X.SaveLLMJson.



To ensure accurate results, you must add all required methods in the correct sequence as shown in the reference screen. Missing or disordered methods may lead to incomplete or incorrect field extraction.

Note: This method **rrSet("-Extract", "@X.SaveLLMJson")** ensures that during the **Extract** phase, the system executes the SaveLLMJson logic, capturing and storing AI-generated results in a structured format for further processing or validation.

The rrset("-Extract", "@X.SaveLLMJson") method in IBM Datacap is used to configure rule execution by setting a specific rule set for the **Extract** action. Here's what each part signifies:

- **Extract:** This is the action or task phase where the rule set is applied—typically during data extraction.

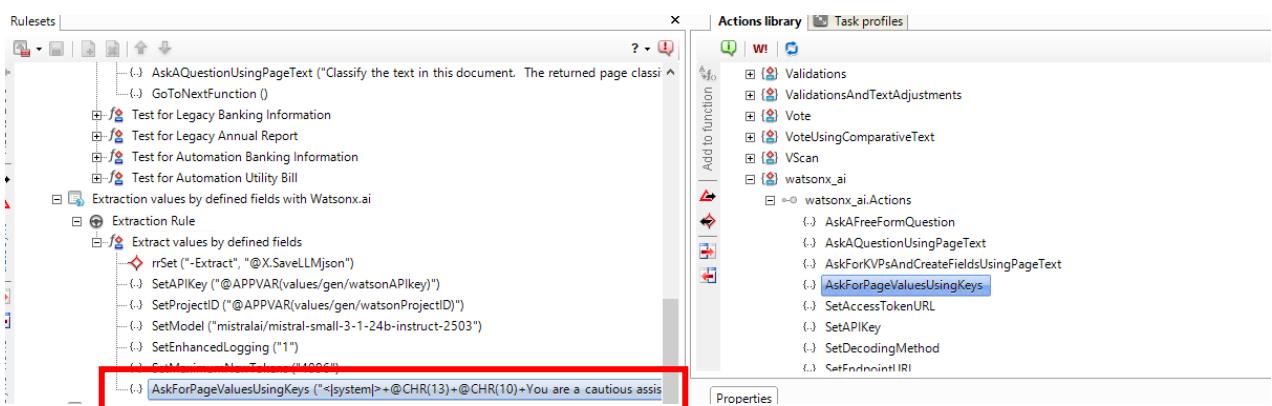
- **@X.SaveLLMJson:** This refers to the rule set or method being assigned. In this case, it likely triggers the SaveLLMJson function from the X ruleset, which is designed to save the output from a large language model (LLM) like Watsonx.ai in JSON format.

■

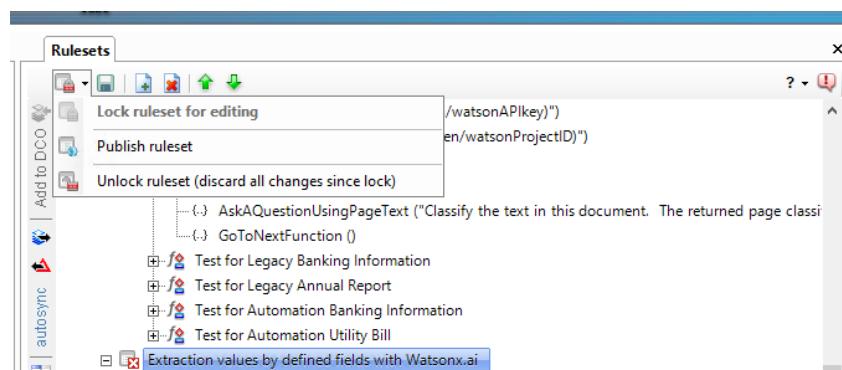
\_ 9. Add the “**AskForPageValuesUsingKeys**” method from watsonx.ai action library to the function and add the

- **String Question** value “<|system|>+@CHR(13)+@CHR(10)+You are a cautious assistant. You carefully follow instructions. You are helpful and harmless and you follow ethical guidelines and promote positive behavior.+@CHR(13)+@CHR(10)+<|user|>+@CHR(13)+@CHR(10)+[Document]+@CHR(13)+@CHR(10)+{{PAGETEXT}}+@CHR(13)+@CHR(10)+[End]+@CHR(13)+@CHR(10)+The following is a list of keys to find within the document. Please find only these keys and do not return any additional keys.+@CHR(13)+@CHR(10)+@CHR(13)+@CHR(10)+{{LLMKEYS}}+@CHR(13)+@CHR(10)+@CHR(13)+@CHR(10)+Return the key value pairs in JSON format. Return only JSON in the following format. Do not include additional keys. Do not include additional text before or after the JSON.+@CHR(13)+@CHR(10)+@CHR(13)+@CHR(10)+<| assistant |> “

- **String format** parameter value to 0



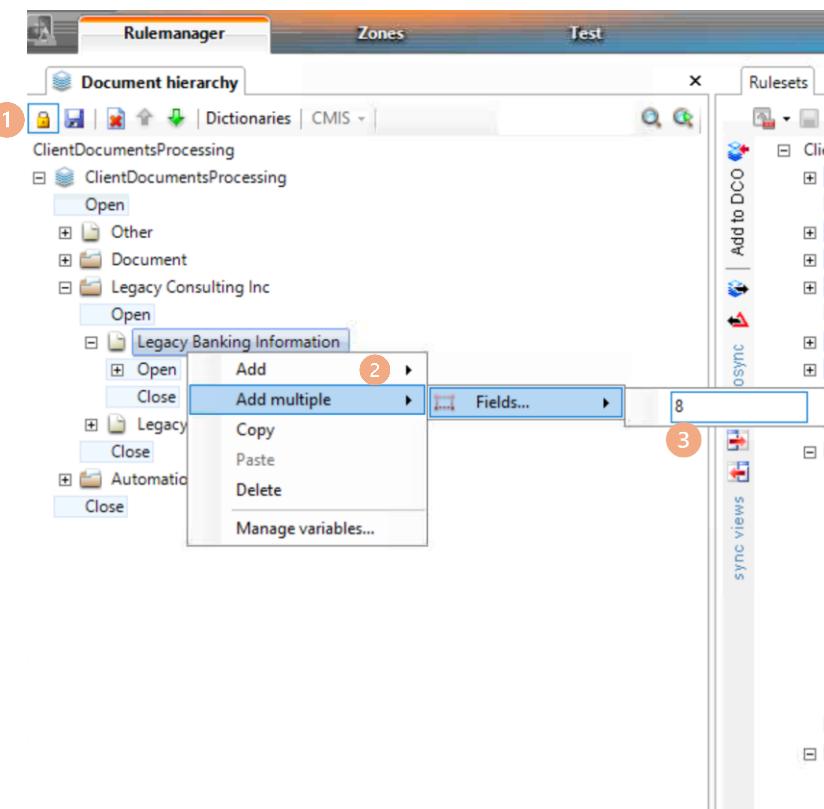
\_ 10. Save and publish the “**Extraction values by defined fields with Watsonx.ai**” ruleset.



■ You have successfully created both the rule sets “*Extraction all fields with Watsonx.ai*” and “*Extraction values by defined fields with Watsonx.ai*” using Datacap Studio.

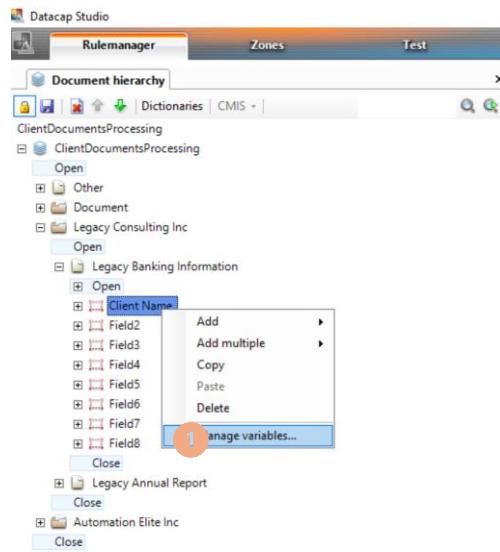
You are going to create the defined fields for the “**Legacy Banking Document**” using **Datacap Studio**, though you can also create these fields using the **FastDoc Admin** application if you prefer.

\_11. Click the **lock icon** in the **Document Hierarchy**, then right-click on **Legacy Banking Information**, select **Add Multiple Fields**, and enter the value **8**, indicating that you will extract **8 fields** from the document.

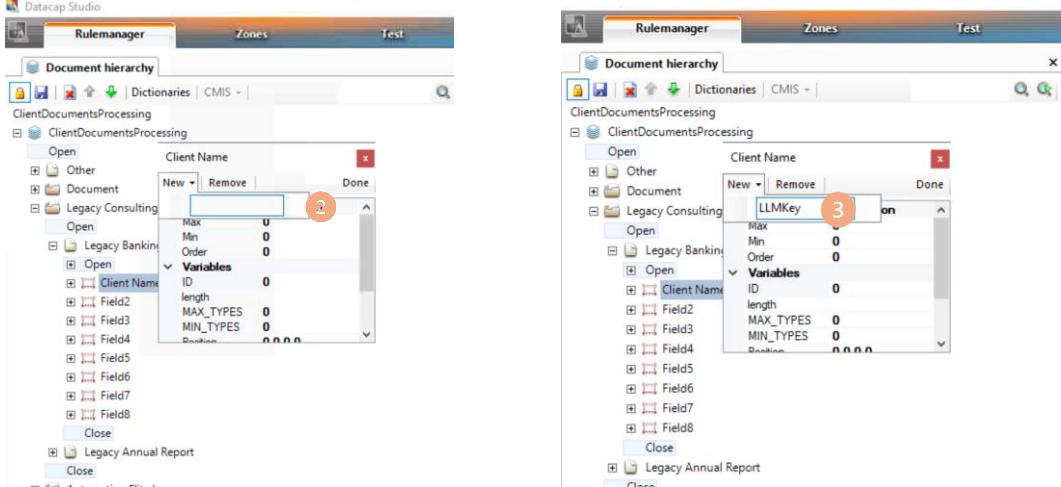


\_12. Rename **Field 1** to **Client Name** from the Properties tab.

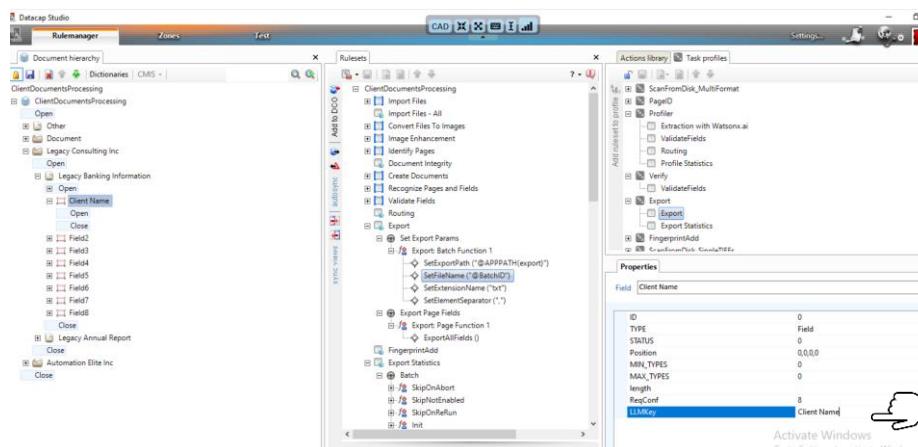
\_13. You need to define the **LLMKey** variable and assign its value to the field by **right-clicking on "Client Name"** and selecting **Manage Variables**.



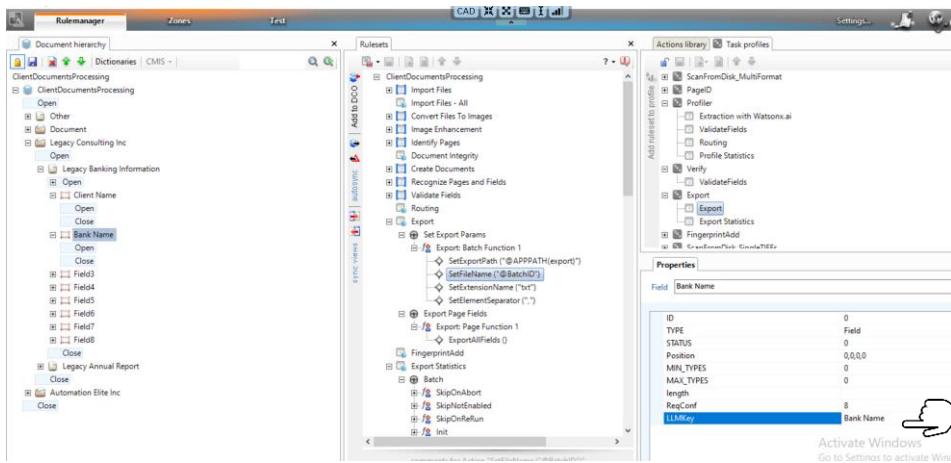
14. Click **New**, add the **LLMKey** variable, and then click **Done** to save the changes.



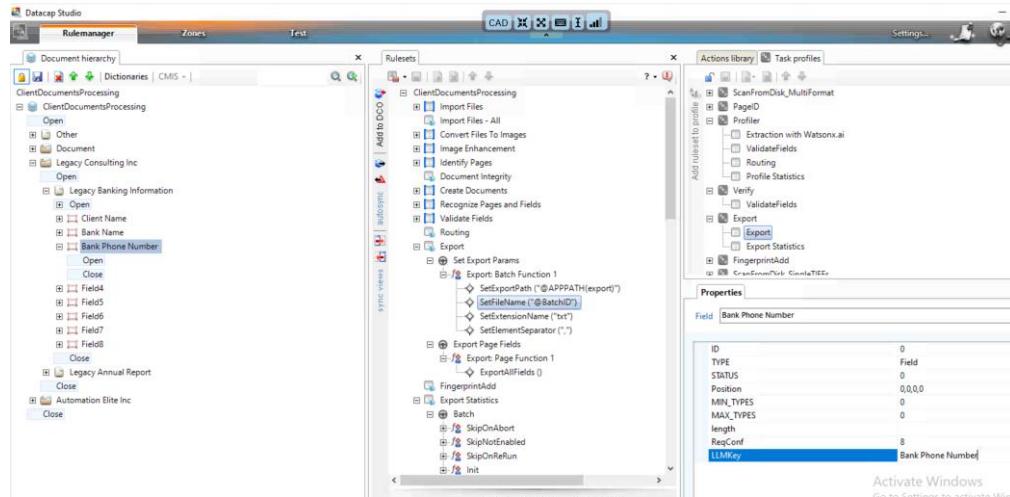
15. Set the **LLMKey** variable value for the **Client Name** field to "Client Name".



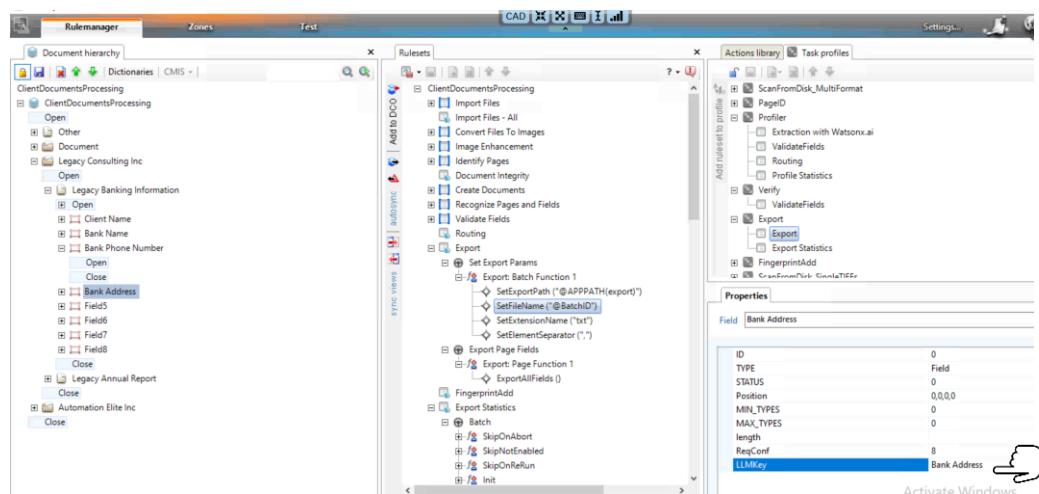
16. Rename **Field 2** to **Bank Name**, then add the **LLMKey** variable and set its value to "**Bank Name**".



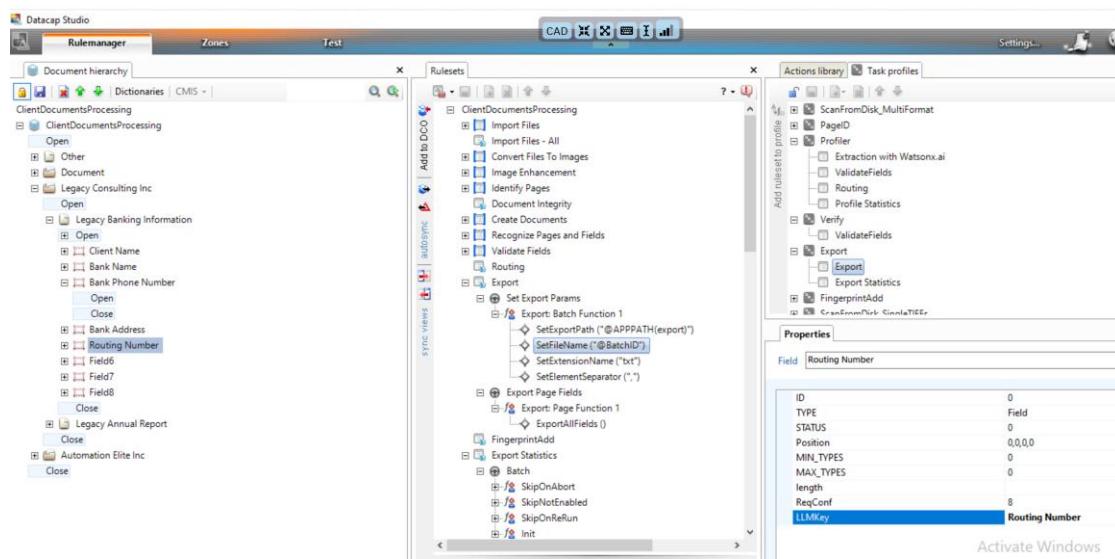
\_17. Rename Field 3 to **Bank Phone Number**, then add the **LLMKey** variable and set its value to "**Bank Phone Number**".



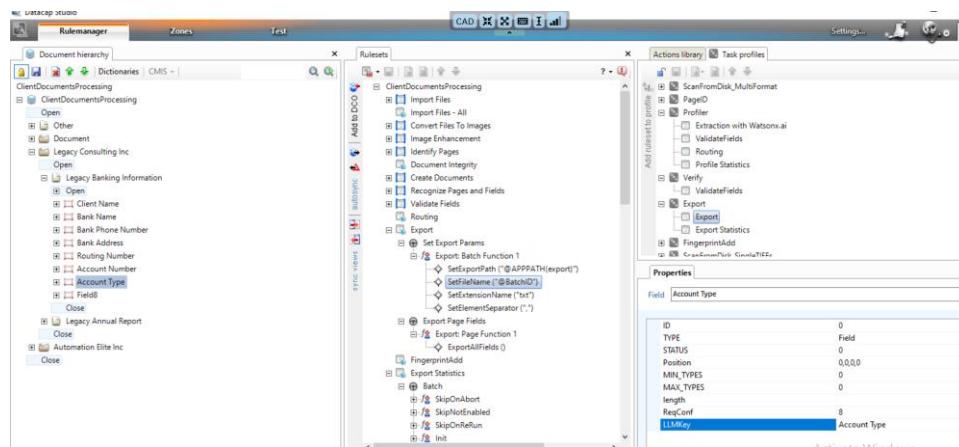
\_18. Rename Field4 to **Bank Address**, then add the **LLMKey** variable and set its value to "**Bank Address**".



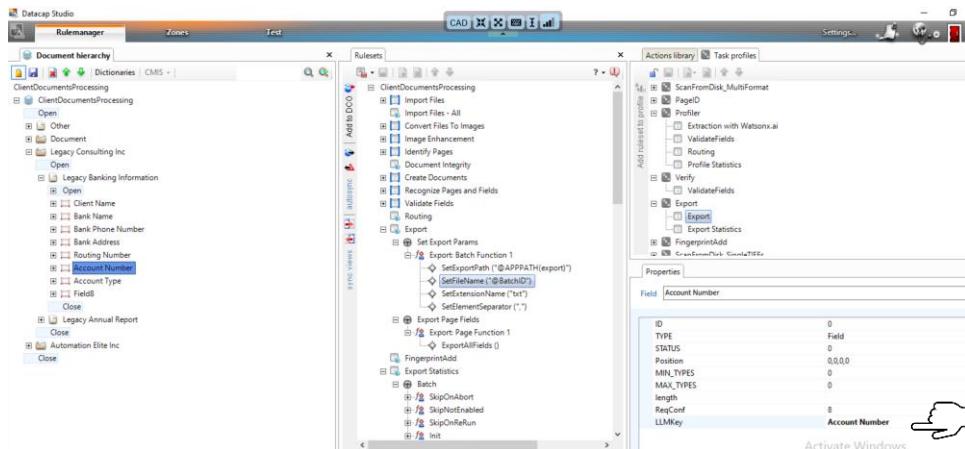
\_19. Rename **Field5** to **Routing Number**, then add the **LLMKey** variable and set its value to " **Routing Number**" .



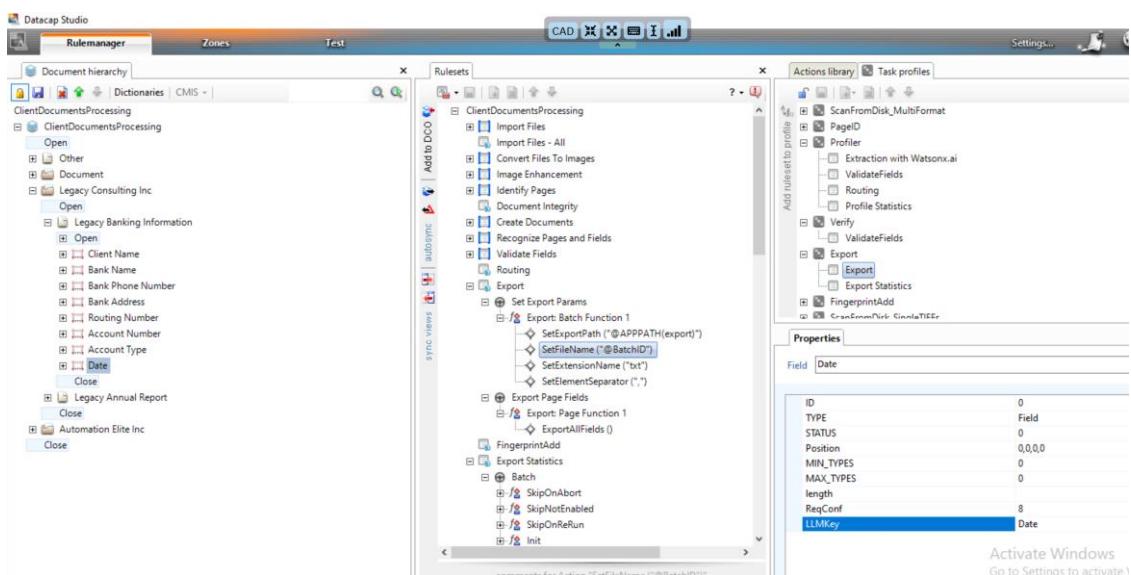
\_20. Rename **Field6** to **Account Type**, then add the **LLMKey** variable and set its value to " **Account Type**" .



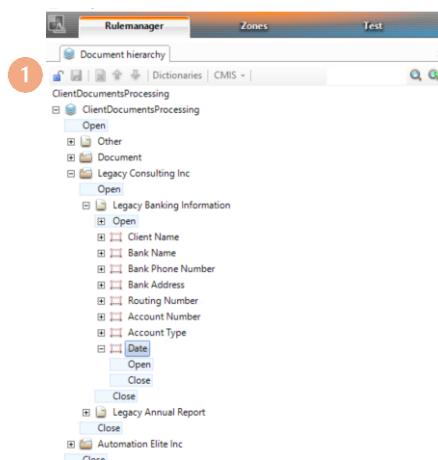
\_21. Rename **Field7** to **Account Number**, then add the **LLMKey** variable and set its value to " **Account Number**" .

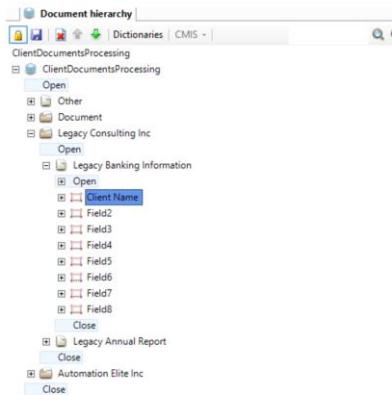


\_22. Rename **Field8** to **Date**, then add the **LLMKey** variable and set its value to “**Date**”.



\_23. Save your changes and click the **lock icon** to unlock the **DCO under Document Hierarchy**, as shown below.





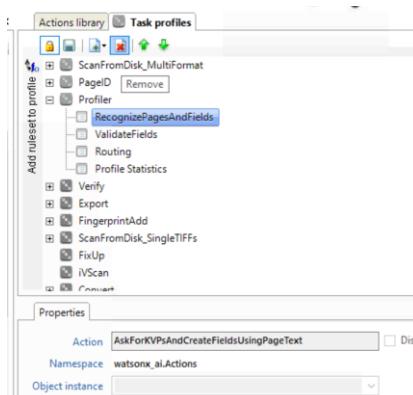
### 3.2.2.5 Map both rulesets “Extraction all fields with Watsonx.ai” and “Extraction values by defined fields with Watsonx.ai” to Task Profile Workflow

\_1. Remove Recognize Pages and Fields ruleset from “**Profiler**” under *Task profiles*. Now add both rulesets “Extraction all fields with Watsonx.ai” and “Extraction values by defined fields with Watsonx.ai” to Profiler task.

Before editing Task Profiles, you must **lock them** by clicking the lock icon on the Task Profiles tab.

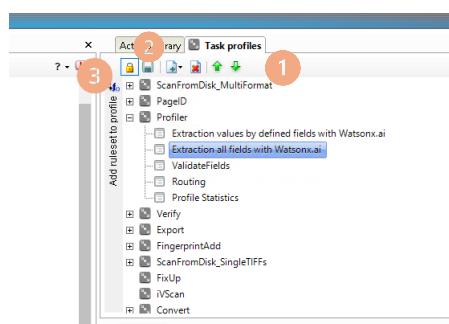
To remove a ruleset, click the “**X**” icon next to the ruleset name.

To add a ruleset, click the “**Add Ruleset to Profile**” button and select the desired ruleset from the list.



\_2. Move both rulesets above validateFields using the arrow icons, click **Save**, then click the **lock icon** to unlock the Task Profiles.

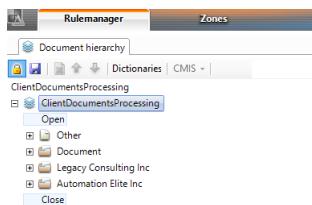
You successfully mapped the extraction rulesets to task profiles.



### 3.2.2.6 Map the Extraction rules to the Document Hierarchy (DCO)

Now you are adding the rules to the DCO (Document Hierarchy) to specify where the ruleset should be executed during processing.

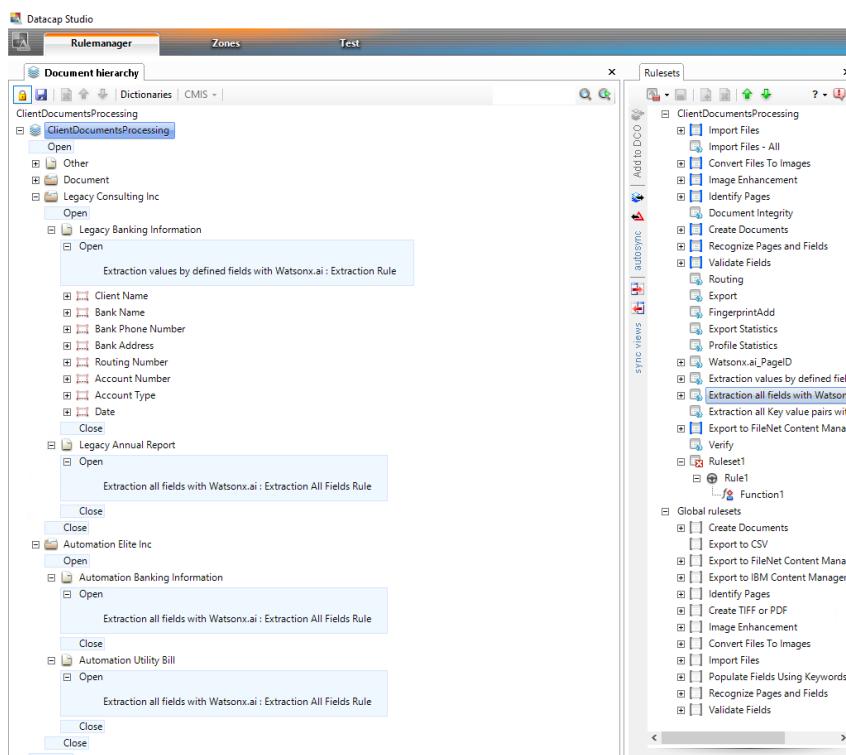
\_1. Lock the Document Hierarchy by clicking the lock icon, then add the extraction Rules from ruleset to the DCO under the Pages by clicking Add to DCO.



\_2. Added the Extraction Rule by defined fields with Watsonx.ai to Legacy Banking Information page Type under the Open Page section, as LLMKeys have been defined for this document and extraction is performed using the AskForPageValuesUsingKeys method with a zero-shot prompt.

\_3. Added the Extraction All Fields with Watsonx.ai rule (Extraction All Fields Rule) to the remaining page types: Legacy Annual Report, Automation Banking Information, and Automation Utility Bill.

After the rules are added to the DCO object, you should see it reflected as shown in the screenshot below.



\_4. Save and click on the lock icon to unlock the DCO.

### 3.2.2.7 Verify/Test the Document Extraction within DataCap studio with Client Documents

In this section you are validating the Extraction rulesets in DataCap with Client Documents.

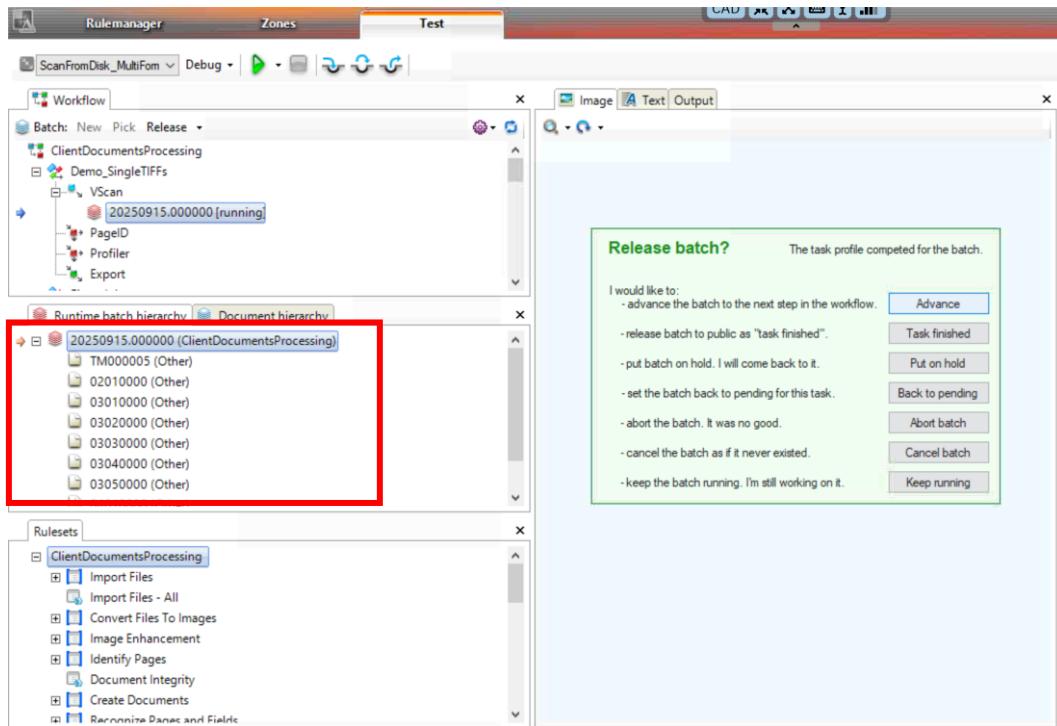
You already placed the documents in the **input folder** during document classification verification section.

\_1. Click **Test** Tab from the Datacap Studio.

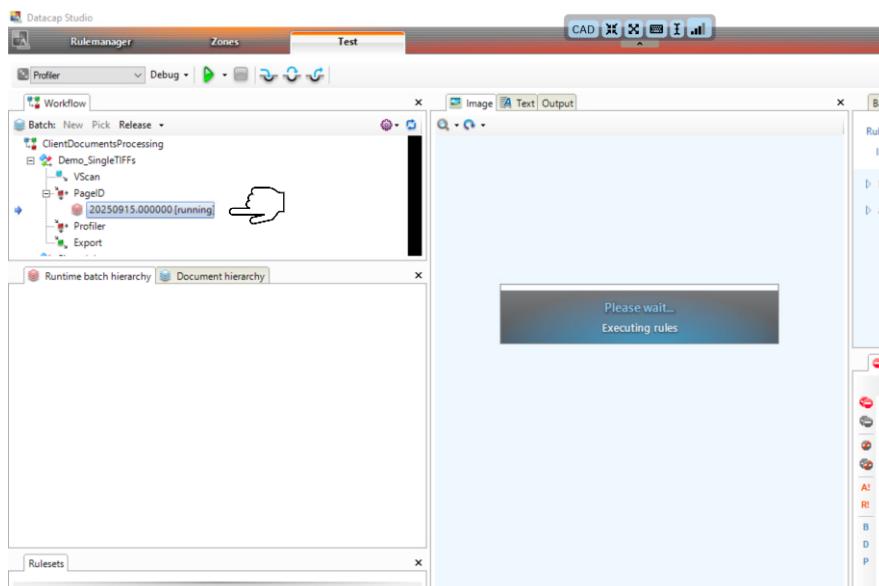
\_2. Select **ScanFromDisk\_MultiFormat** from the dropdown menu as shown below.

\_3. Right click on the Vscan under **Demo\_singleTiffs**, click New and click the green **Play** button to start the process.

\_4. Click **Advance** button to move to the **PageID** task once it's scanned and convert pdf files into tiff files.

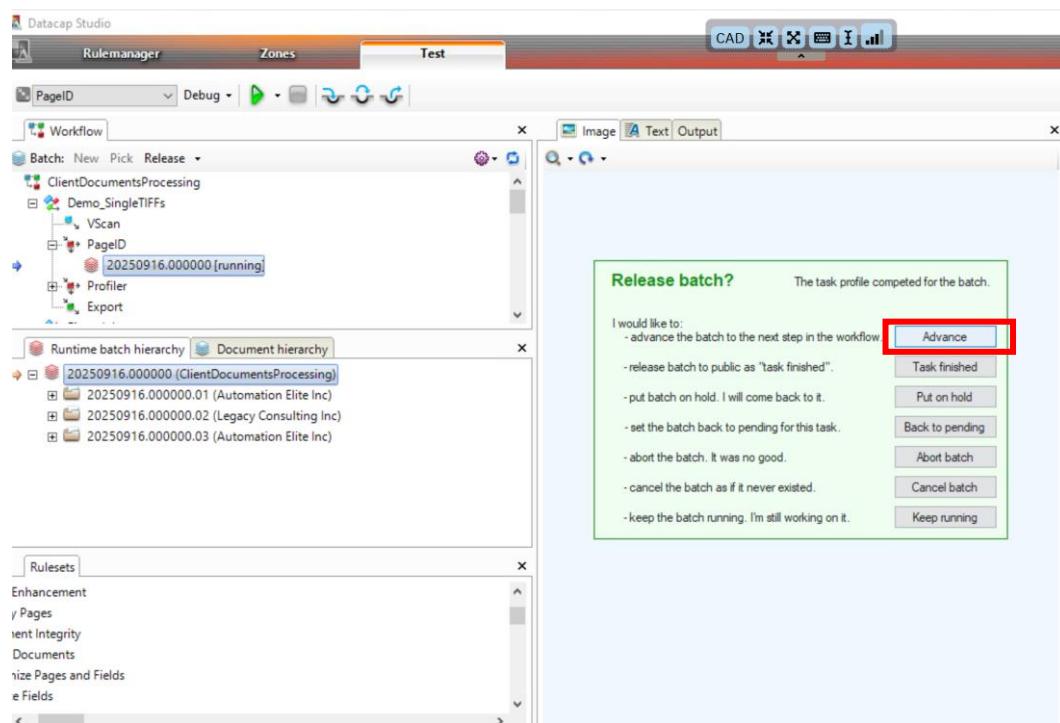


\_5. Now **PageID** ruleset is processing for recognition and classification of the documents.



\_6. Once the PageID is successfully completed you can see the documents get **classified**. Click Advance to move from PageID task to Profiler where we are extracting the fields from the documents.

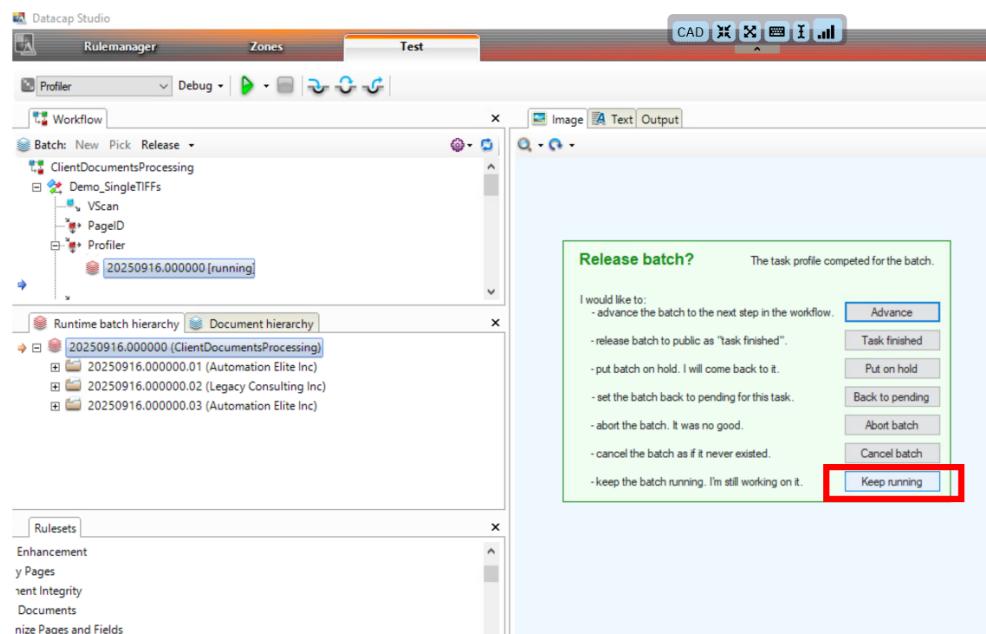
**Note:** you must click the green **Play** button to start the process.



- We will be covering the following steps for completing the Profiler task:

## 7. Profiler Task Completion:

- Once the Profiler task is successfully completed, click **Keep running** to keep the process active for the moment.



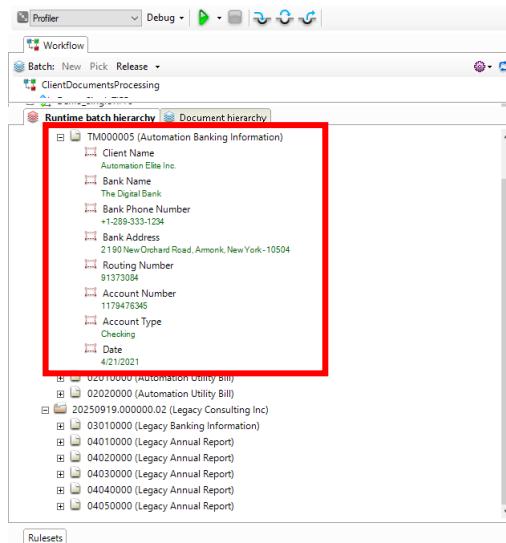
## 2. Create Verify Ruleset:

- Open the Navigator Desktop.
- Create a verify ruleset to validate the extracted fields and document.
- This ruleset will be used to verify that the fields extracted are accurate and the document complies with your validation criteria.

## 3. Export Document with Fields:

- After verification, use the **Export to FileNet** connector.
- This exports the document along with its verified fields into your FileNet repository.

\_8. You can view the extracted fields by expanding each **document folder** in the **Runtime Batch Hierarchy** to verify the results.



\_9. You can open the **processed batch** from path C:\Datacap\ClientDocumentsProcessing\batch and view the extract.json file, which contains the **Legacy Banking Information** response returned by **Watsonx.ai LLM** (for today's date and running batch). You may view the same from the studio as well where you will find it in Document Hierarchy.

This PC > Local Disk (C:) > Datacap > ClientDocumentsProcessing > batches > 20250916.000000				
	Name	Date modified	Type	Size
	02010000.cco	9/16/2025 8:11 PM	CCO File	17 KB
	02010000.tif	9/16/2025 8:09 PM	TIF File	11 KB
	02010000.txt	9/16/2025 8:11 PM	Text Document	1 KB
	02010000.xml	9/16/2025 8:14 PM	XML Document	10 KB
	02010000_layout.xml	9/16/2025 8:11 PM	XML Document	63 KB
	02010000c.xml	9/16/2025 8:11 PM	XML Document	13 KB
	02010000-Extract.json		JSON File	1 KB
	03010000.cco		CCO File	37 KB
	03010000.tif		TIF File	26 KB

```

1  {
2      "Client Name": "Legacy Consulting",
3      "Bank Name": "XYZ Bank",
4      "Bank Phone Number": "+1-213-111-7890",
5      "Bank Address": "887 Cypress Rd, Garden Grove, CA 92840",
6      "Routing Number": "32043559",
7      "Account Number": "7250512345",
8      "Account Type": "Checking",
9      "Date": "5/10/21"
10 }

```

### 3.2.2.8 Document verification

Create verify Ruleset(Rules, and Functions) for the ClientDocumentsProcessing app with Datacap Studio

In this section, you will create a verify ruleset with the necessary rules and functions to validate the extracted document fields before exporting the document and its properties and saving them to FileNet.

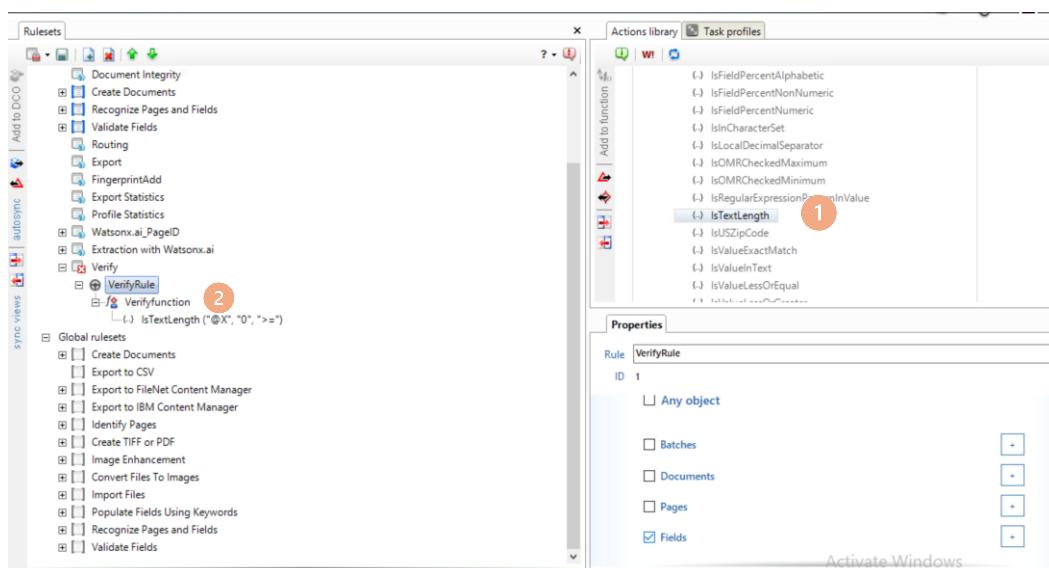
\_1. Add another new ruleset to ClientDocumentsProcessing application

\_2. Rename **Ruleset1**, **Rule1**, and **Function1** to **Verify**, **VerifyRule**, and **VerifyFunction** respectively, as shown below.



\_3. Add **IsTextLength** from *ValidationsAndTextAdjustments* action library and add the following parameters

- String Target = "@X"
- String Length="0" and
- String ComparisonOperator= ">="



\_4. Click on the **Verify Rule** and check the **Fields** value under 'Run rule at the start of ..' from the **Properties** panel on the right as shown below:

\_5. Save and publish the verify ruleset.

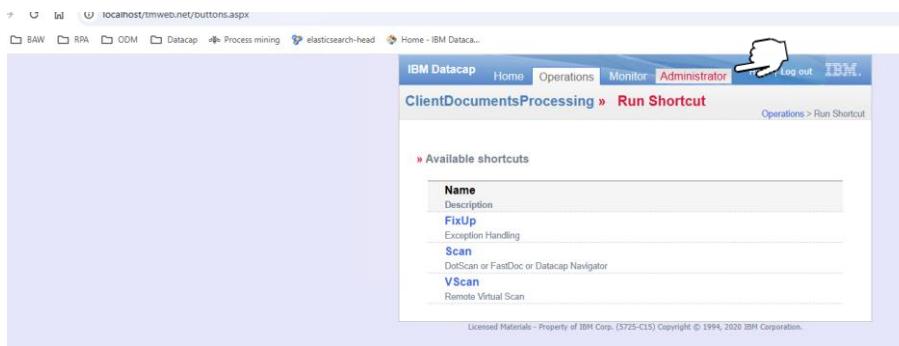
### 3.2.2.9 Map verify Ruleset to Task Profile Workflow

Before editing Task Profiles, you must lock them by clicking the lock icon on the Task Profiles tab.

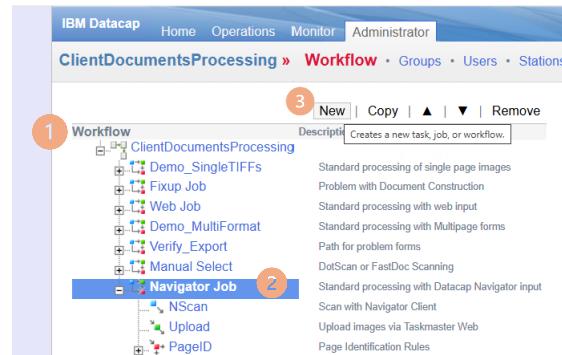
If verify task is not available in the task profiles, you can create it from tmweb url <http://localhost/tmweb.net/> or you can also do it later from navigator desktop exercise. [Skip this section for the moment.](#)

\_1. Login the **tmweb.net** by selecting **ClientDocumentsProcessing** with login credentials **admin/admin** and **station 1**.

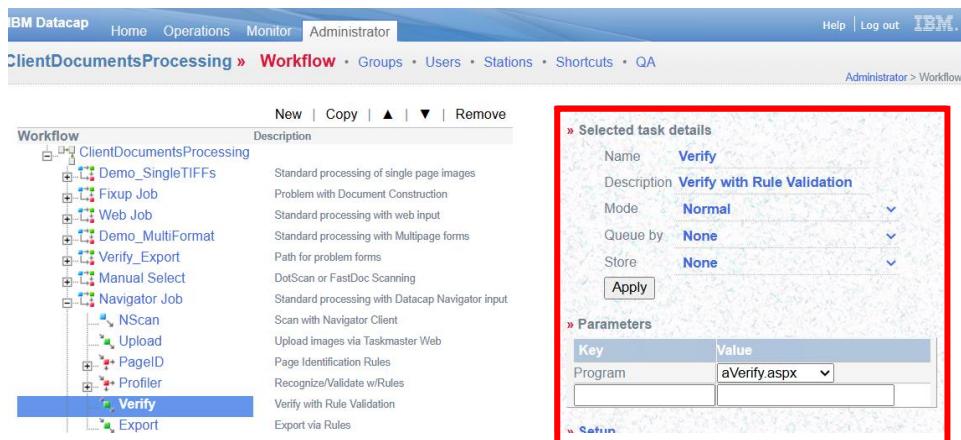
\_2. Click on the **Administrator** Tab



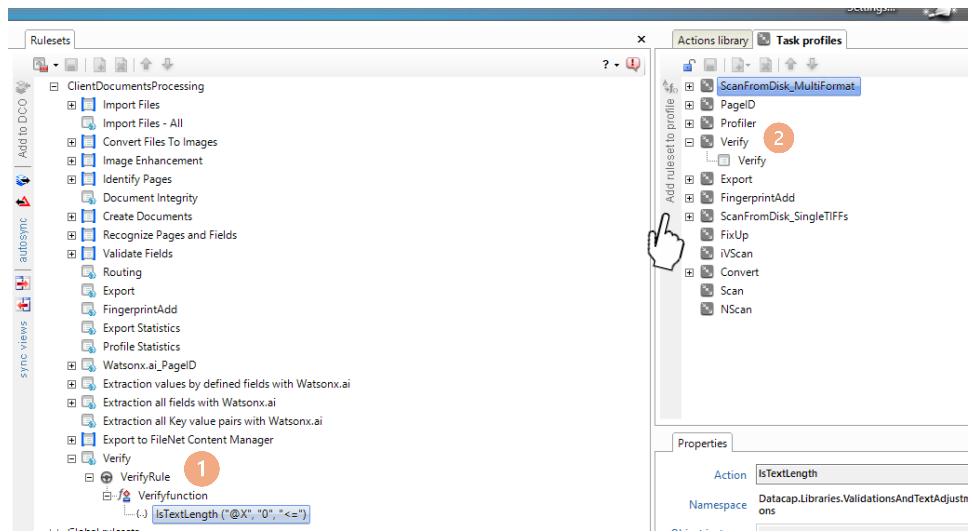
\_3. Click on the **workflow** on the left panel and select the Navigator Job where you want to add to the workflow task profiles. For the same, click on **New**.



\_4. Give the **task details** (*Name, Description, Mode and Program*) as shown below and click **Apply**.



\_5. You can now click the **lock icon** and map the verify ruleset to the verify task in the task profiles by clicking on **Add ruleset to profile**.

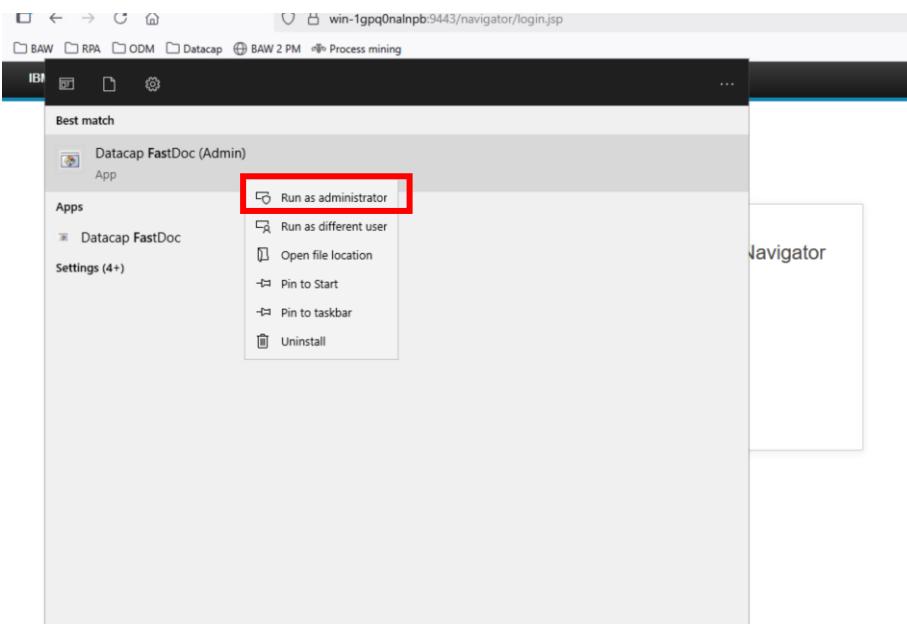


\_6. Save the changes, and then click the lock icon again to unlock the task profiles.

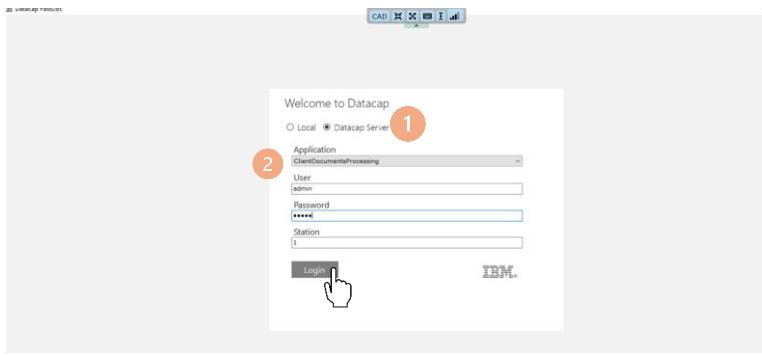
### 3.2.2.10 Export configuration to FileNet content Manager repository

In this section, you will configure **FileNet Content Manager** to **export client documents** after they have been processed using **Datacap** and **Watsonx.ai**.

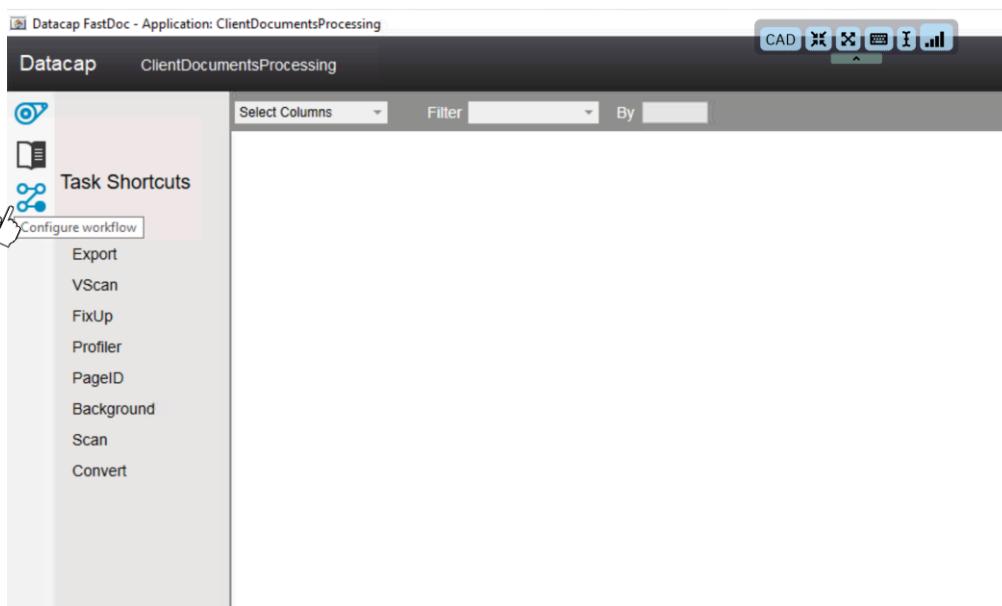
\_1. Open the **Datacap FastDoc Admin application** from the **Windows Start menu**, Run as administrator.



\_2. Select the **Datacap Server**, choose the **ClientDocumentsProcessing** application, and set the **Station ID** to **1**. Log in using the credentials **admin/admin**.



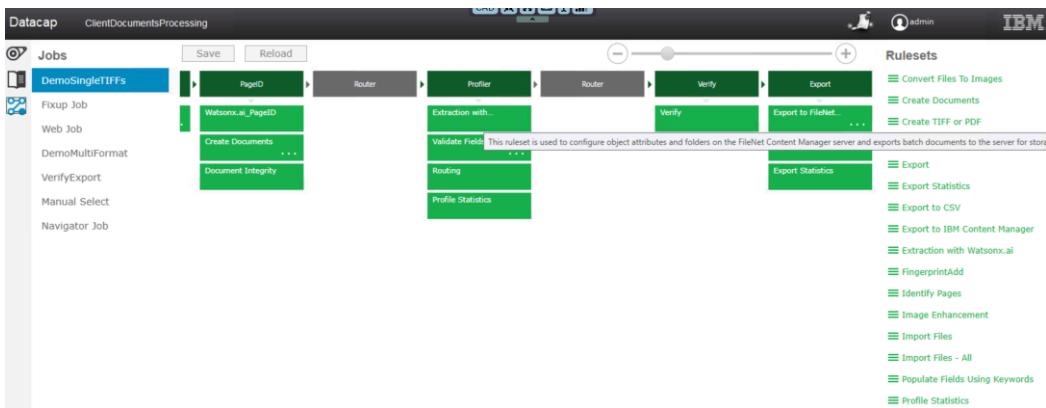
\_3. Click on the **Workflow** icon  [Configure workflow](#) to view the existing workflow configured for the **ClientDocumentsProcessing** application.



\_4. Drag and drop the **Export to FileNet Content Manager** ruleset from the **Rulesets** panel to the **Export** task profile.



\_5. You will now see **Export to FileNet** listed under the **Export** task profile.

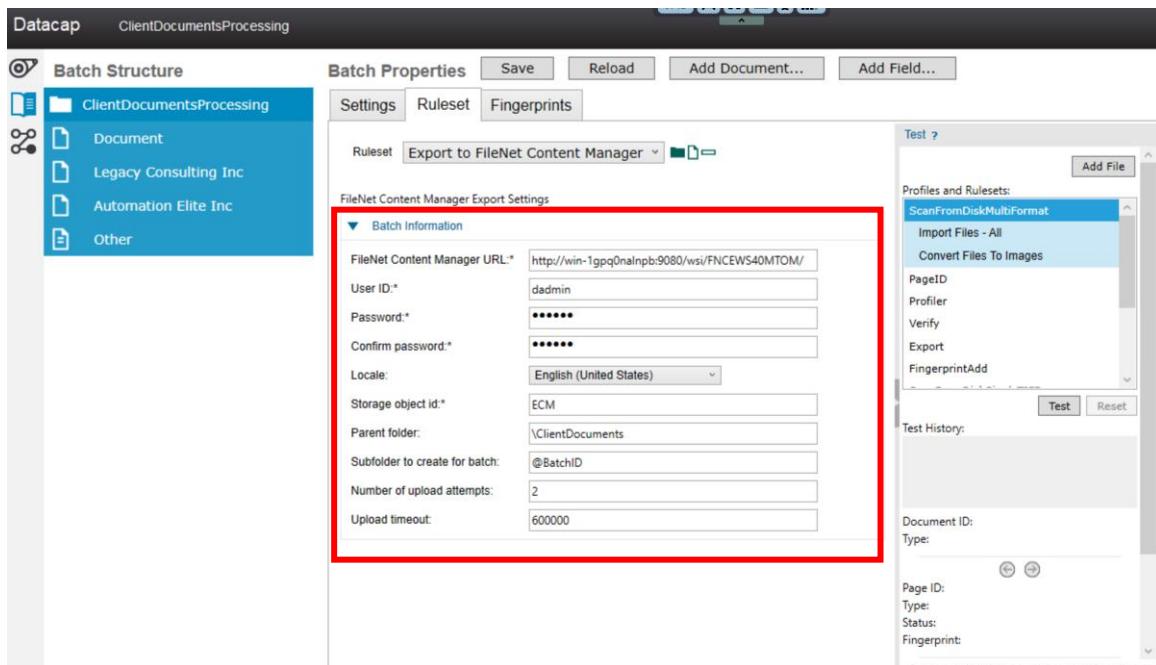


\_6. If prompted to save, simply click **Save** to confirm and proceed.

\_7. Double-click on **Export to FileNet**, and you will see the **FileNet connection settings** window open.

\_8. Provide the **FileNet connection settings** including *URL*, *UserID* (*dadmin*), *Password* (*dadmin*), *Object Store*, *Folder*, *Subfolder*, and *Upload Timeout* as shown below.

URL <http://win-1gpq0nalnpb:9080/wsi/FNCEWS40MTOM/>



\_9. To proceed, click on **Legacy Consulting Inc** under the **Batch Structure** section in Datacap FastDoc Admin. This will allow you to configure and manage the batch structure for this specific client, including defining document types, pages, and associated rulesets.

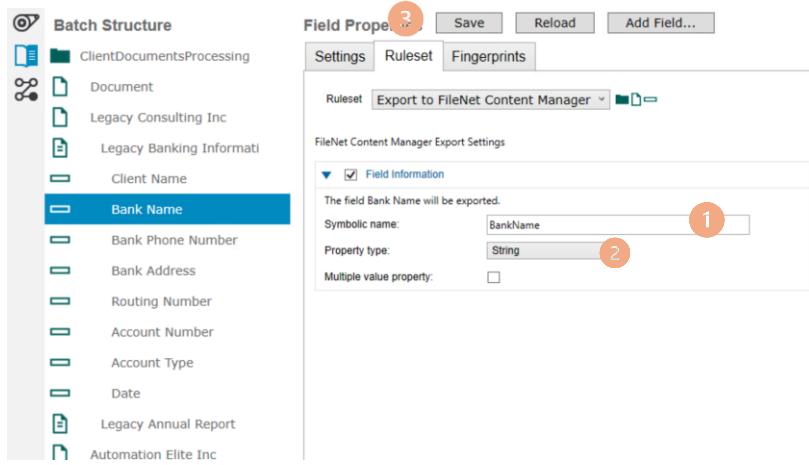
- Provide the document title, document class ID, and file extension details for the Legacy Consulting Inc document class within the batch structure as illustrated below.

You will create a document class in IBM FileNet with the same name and properties as defined in the Datacap application.

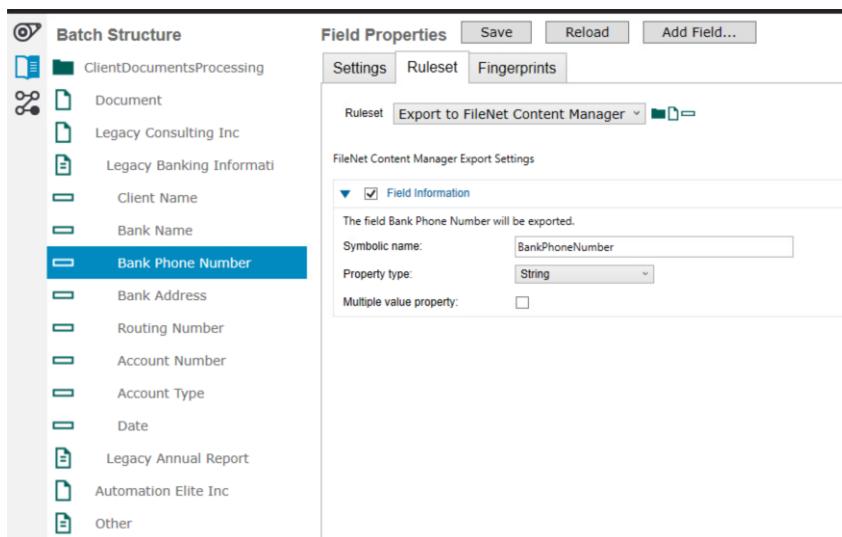
\_10. Expand the Legacy Banking information section in the left Panel of documents under Legacy Consulting Inc until you reach the **fields** related to **Legacy Banking information**. Then, select the '**Client Name**' field and provide the Ruleset information as demonstrated below.

- check the Field Information and note the symbolic name as '**ClientName**' and the property type as '**String**'

\_11. Select the **Bank Name** field and provide the Ruleset information as demonstrated below. check the Field Information and note the symbolic name as '**BankName**' and the property type as '**String**'.



\_12. Select the **Bank Phone Number** field and provide the Ruleset information as demonstrated below. check the Field Information and note the symbolic name as '**BankPhoneNumber**' and the property type as '**String**'

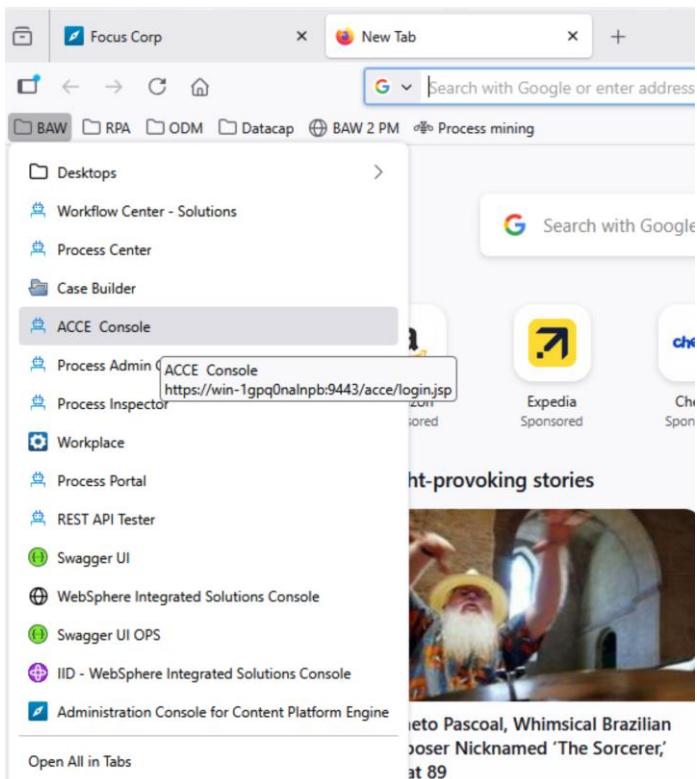


You've mapped three fields so far due to time constraints, but it's entirely up to you—feel free to continue the same process for all remaining fields to ensure thorough and consistent mapping across your dataset.

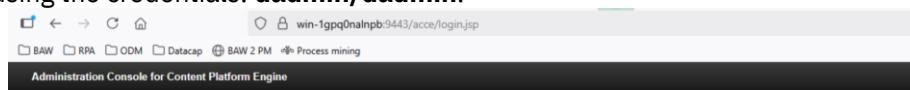
You have successfully configured the FileNet connection, document and field details, and added the Export to FileNet Ruleset to the Export Task profiles.

Next you are going to configure the Document Class and its properties in the FileNet content repository using the ACCE console.

\_1. Log in to the ACCE console using the BAW bookmarked favorite in Firefox, then click on **ACCE Console**.



Log in using the credentials: **dadmin/dadmin**.



Welcome to Administration Console for Content Platform Engine

User Name

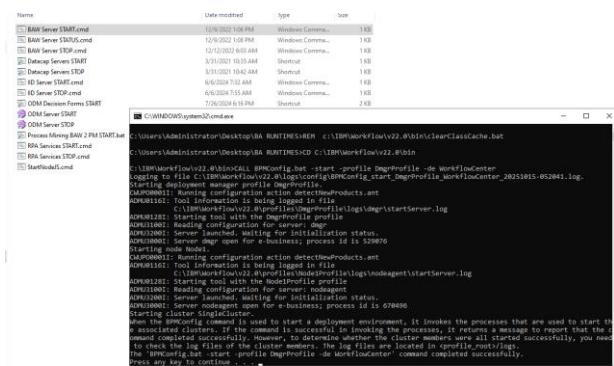
Password

(c) Copyright 2013 IBM Corp. Licensed Materials - Property of IBM. IBM and the IBM logo are trademarks or registered trademarks of IBM Corp. in the U.S., other countries, or both. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

If you're unable to access the ACCE Console and encounter a message indicating restricted access or service issues, refer to **point no. 2** in the instructions to initiate the **WebSphere services** and restore functionality, otherwise continue from **point no 3**.



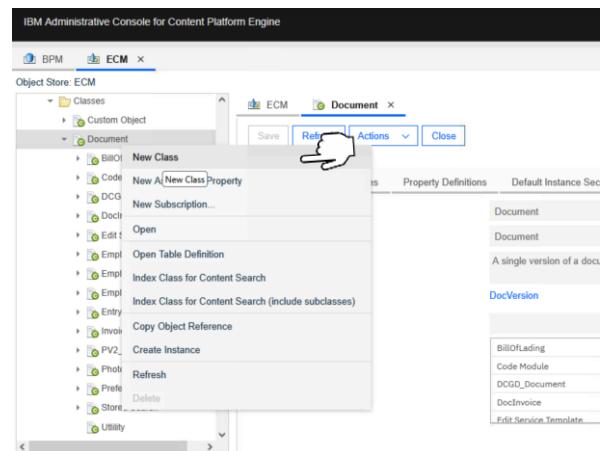
Navigate to your **Desktop** and open the folder named **BA Runtimes**. Inside the folder, run the file titled “**BAW Server START.cmd**”. Once the services have successfully started, you will see the following confirmation message displayed on your screen.



**Note:** It might take few minutes to start this service.

\_2. Click on **ECM Object Store** from the left-hand menu.

\_3. In the left-hand menu, under **Classes**, click on the Document folder, right-click it, and select **Add New Class**.



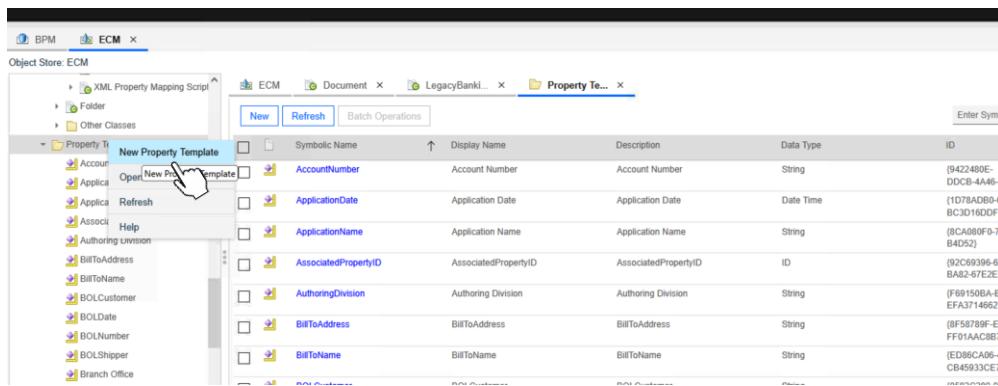
Add the **Display Name** and **Symbolic Name**. The Symbolic Name should match the one you provided while configuring the FileNet connection from the FastDoc App.

Display name: <input type="text"/>	LegacyBankingInformation
Symbolic name: <input type="text"/>	LegacyBankingInformation
Description: <input type="text"/>	LegacyBankingInformation

\_4. Click **Next**, then **Next** again, and finally click **Finish**.

Name	Value
Display name	LegacyBankingInformation
Symbolic name	LegacyBankingInformation
Description	LegacyBankingInformation

\_5. Right-click on **Property Templates** in the left-hand menu and select **New Property Template**.



\_6. Enter **ClientName** as both the *Display Name* and *Symbolic Name* and **Enter** and click **Next**. Ensure that the symbolic names match those used in your previous steps during the **Export to FileNet configuration** from the FastDoc app.

Display name:	Symbolic name:
ClientName	ClientName

\_7. Enter the **Data Type** as **String** and click on **Next**.

Data type:
String

\_8. Select 'Single' value and click **Next**.

Choose whether the object property can hold a single value or multiple values. Most properties are single value, but if the property will contain a list, then choose multi-value.

Single (1)

Multi (2)

List order:

- Non-unique and ordered values (such as lines in address) (3)
- Unique and ordered values (such as the list of primary colors) (4)

Set other attributes (5)

Click **Finish** once you have verified all the details.

Name	Value
Display name	ClientName
Symbolic name	ClientName
Description	ClientName
Data type	String
Assign choice list	False
Assign marking set	False
Single or multi-value	Single
List order	Non-unique and ordered values (such as lines in address)
Set other attributes	False

Create the remaining properties such as **BankName**, **BankPhoneNumber**, and others as needed.

\* Display name: (1)

Existing names:

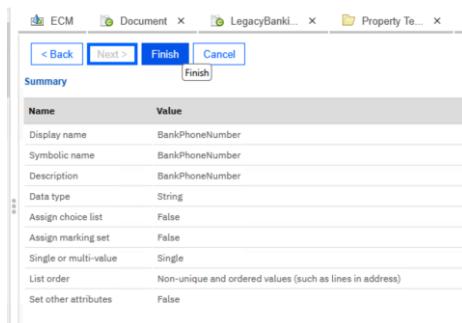
- Account Number
- AccountNumber
- Activity Stream Extended Settings
- Activity Stream Gadget URL
- Activity Stream Generation Enabled
- Activity Stream HTTP Endpoint URL
- Activity Stream Ignored User Ids
- Activity Stream Retrieval URL
- Anonymous Download Count
- App Id

\* Symbolic name: (2)

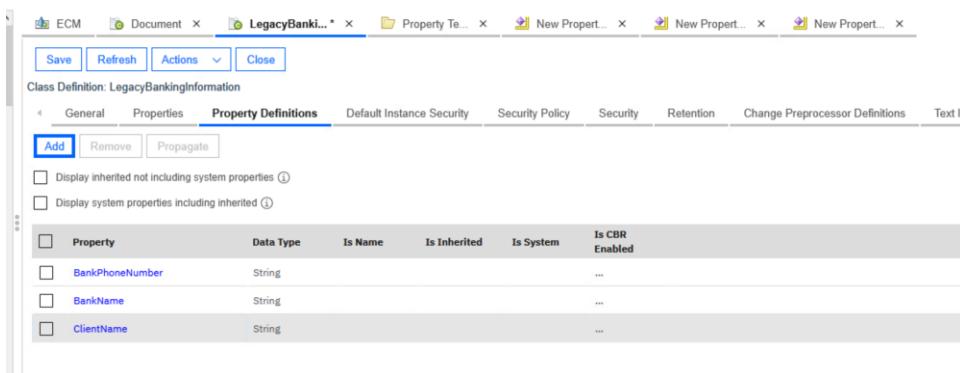
Description: (3)

Activate Wi  
Go to Settings t

Name	Value
Display name	BankName
Symbolic name	BankName
Description	BankName
Data type	String
Assign choice list	False
Assign marking set	False
Single or multi-value	Single
List order	Non-unique and ordered values (such as lines in address)
Set other attributes	False

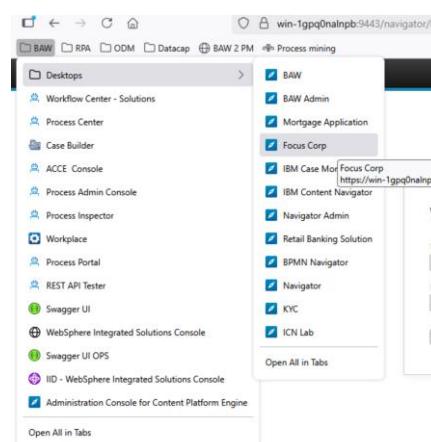


\_9. Map the newly created properties to **LegacyBankingInformation** by clicking on **Property Definitions**, adding the newly created properties, and then saving the document class as shown below.

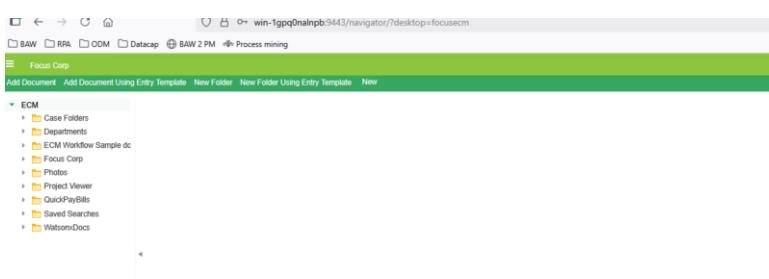


Now, you are going to create a folder under **ECM** to export the Datacap-processed batches into the **ECM repository** within **FileNet Content Manager**.

\_10. Log in to the **Focus Crop** desktop by navigating through the **BAW bookmark** → **Desktops** → **Focus Crop**.



Login with **dadmin/dadmin**.



11. Click on New Folder.

The screenshot shows the ECM interface with a sidebar containing links like Case Folders, Departments, ECM Workflow Sample dc, Focus Corp, Photos, Project Viewer, QuickPayBills, Saved Searches, and WatsonDocs. The main area displays a table of folders with columns for Name, Modified By, and Modified On. A cursor is hovering over the 'New Folder' button in the top navigation bar.

Enter the **Name** as “**ClientDocuments**”, matching the name used earlier during the **Export to FileNet** configuration in the **FastDoc App**.

New Folder  
The values that you enter for the folder properties can be used to find the folder later.

General

\* Save in: ECM

Properties

\* Class: Folder

\*Folder Name: ? ClientDocuments

Activate Windows Go to Settings Cancel Add

You have successfully configured **FileNet** and **Datacap** to export processed documents using **Datacap + watsonx.ai**.

### 3.3 Exercise 3: Datacap Navigator (optional)

For this exercise you will require the following IDs and passwords:

Tool	User ID	Password
IBM Content Navigator	dadmin	dadmin
IBM Datacap Navigator	admin	admin

#### 3.3.1 Datacap Navigator System Overview

Business Automation Navigator provides a user interface for Datacap. It is bundled in CloudPak for Business Automation. The user accesses the system using a web browser.

Datacap includes a plugin component that operates within Content Navigator. **Content Navigator** accesses and executes the core Datacap services using the *Datacap Windows Server* (formerly wTM). Datacap Windows Service runs as a Microsoft Windows service that presents RESTful web services endpoints that expose the Datacap capabilities.

Datacap uses its **Rulerunner service** to execute image processing in background, on servers. Rulerunner executes image cleanup, conversion, OCR, validation rules and other functions.

Datacap stores work-in-process images and data files on a File System and process tracking data in a database. Users are authenticated using LDAP or Active Directory.

Completed documents are stored in a *FileNet Content Management repository*.

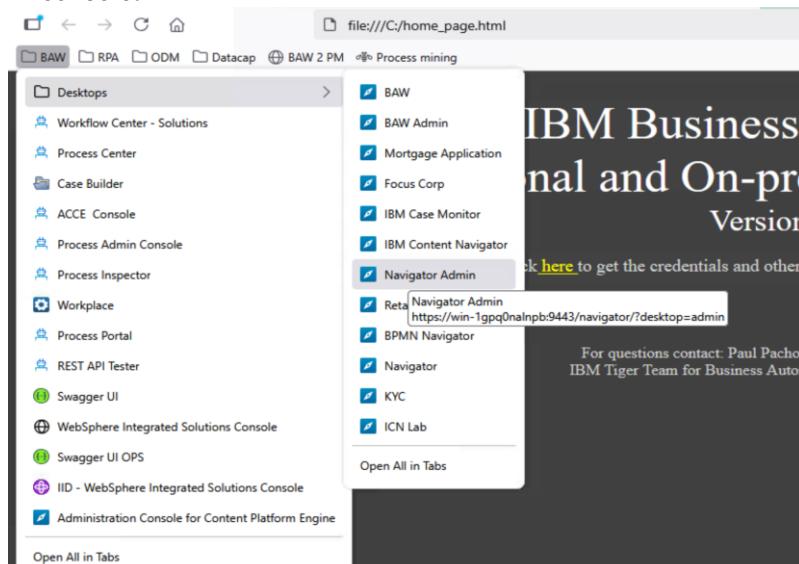
#### 3.3.2 Datacap Navigator Introduction – Step by Step Instructions

To get started, you first need to create Datacap repository and desktop that uses the **ClientDocumentsProcessing application** which you created in the previous exercises.

To configure on IBM Navigator execute the following steps:

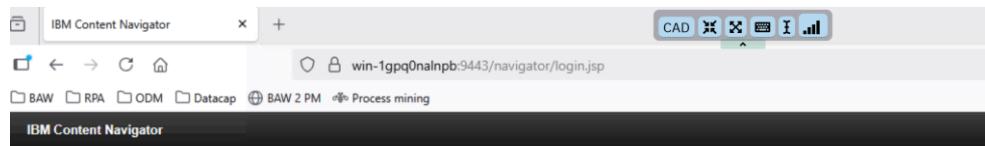
Open **Chrome** or **Firefox** on the VM, where all the required **URLs** and **user credentials** are pre-configured and **bookmarked** for easy access.

\_1. Click the **BAW Bookmarks** (Favorites) folder, then select **Desktops**, and click on **Navigator Admin** to open the admin console.



\_2. Login to the **Navigator Admin** desktop using the provided **credentials** to access the administration console.

**Credentials:-** dadmin/dadmin



\_3. Select '**Repositories**' from the left-hand side.

\_4. Click on New Repository and select '**Datacap Application**' from drop down.

\_5. Enter the following configuration parameters:

A screenshot of the "New Repository" configuration page. The top navigation bar shows "Desktops", "Repositories", and "\*New Repository". The "Save and..." button is highlighted with a red circle labeled 1. The "Save", "Reset", and "Close" buttons are also visible. The main form has tabs for "General" and "Configuration Parameters". Under "General", the "Display name" field contains "ClientDocumentsProcessing" (circled 1). The "ID" field contains "ClientDocumentsProcessing" (circled 2). The "Datacap wTM URI" field contains "http://localhost:8070/ServiceWTM.svc" (circled 3). The "Application" dropdown menu is open, showing options like "GoldenDemo", "FormTemplate", "LearningTemplate", "Transaction", "WatsonaiRedaction", and "ClientDocumentsProcessing" (circled 4). The "Default Station" and "Use ActiveX in IE" fields are collapsed. The "Use Virtual Viewer" field is expanded, showing options like "ClientDocumentsProcessing". A "Connect..." button is at the bottom.

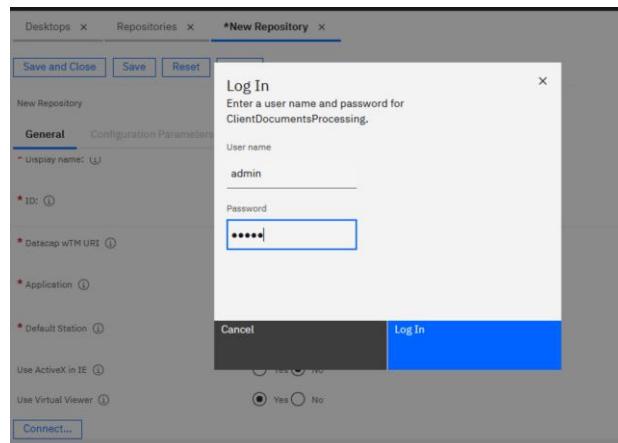
- The '**Display name**' is what we are going to call our new repository. This is not creating a “true” repository, but simply creating a pointer to an application. By putting a “dc” in front, it helps administratively to distinguish Datacap application from actual repositories.
- The “**ID**” is automatically set for you.
- The “**Datacap wTM URL**” is the address to the **wTM service** (which presents RESTful web services endpoints that expose the Datacap capabilities).

- For the “**Application**”, select the “**ClientDocumentsProcessing**” you created earlier in previous exercises. Note it may take a second or two for the wTM service to recognize. If you do not see your application listed, verify your Datacap wTM URL is correct.
- Select **No** for “**Use ActiveX in IE**” and **Yes** for “**Use Virtual Viewer**”.

The screenshot shows the 'New Repository' configuration dialog with the 'General' tab selected. The 'Display name' field contains 'ClientDocumentsProcessing'. The 'ID' field contains 'ClientDocumentsProcessing'. The 'Datacap wTM URI' field contains 'http://localhost:8070/ServiceWTM.svc'. The 'Application' dropdown is set to 'ClientDocumentsProcessing'. The 'Default Station' dropdown is set to '1'. Under 'Use ActiveX in IE', the 'No' radio button is selected (marked with a red circle containing the number 5). Under 'Use Virtual Viewer', the 'Yes' radio button is selected (marked with a red circle containing the number 6). A 'Connect...' button is at the bottom.

\_6. Click **Connect**.

\_7. Login to the Datacap server with user **admin** and password **admin** and click on Log In.



You can leave all the default columns to display to the users in job monitor under configuration parameters. These selected columns can later be edited in the user settings.

Desktops X   Repositories X   \*New Repository X

**Save and Close** **Save** **Reset** **Close**

New Repository

**General** Configuration Parameters

* Display name: ⓘ	ClientDocumentsProcessing
* ID: ⓘ	ClientDocumentsProcessing
* Datacap wTM URI ⓘ	http://localhost:8070/ServiceWTM.svc
* Application ⓘ	ClientDocumentsProcessing
* Default Station ⓘ	1
Use ActiveX in IE ⓘ	<input type="radio"/> Yes <input checked="" type="radio"/> No
Use Virtual Viewer ⓘ	<input checked="" type="radio"/> Yes <input type="radio"/> No

\_8. Click **Save and Close**.

You can either **copy** an existing desktop or **create a new one**; the steps below demonstrate how to **copy a desktop** from the existing **Datacap desktop** configuration.

\_9. Click on “**Desktops**” on the left-hand side.

\_10. Right-click on the existing “**Desktop**” with the **dcgd** ID and select **copy**.

Desktops 1   Repositories   Sync Services   FileNet Content Manager   Content Manager OnDemand   Viewer Maps   Plug-ins   Menus   Labels   Themes   Icon Mapping   Settings   Role-based Redactions   Role-based Desktop Administration

Desktops X   Repositories X

You can create multiple desktops to give different users access to the content they need. For example, you can create a desktop that only allows users to search for content or a desktop that gives user

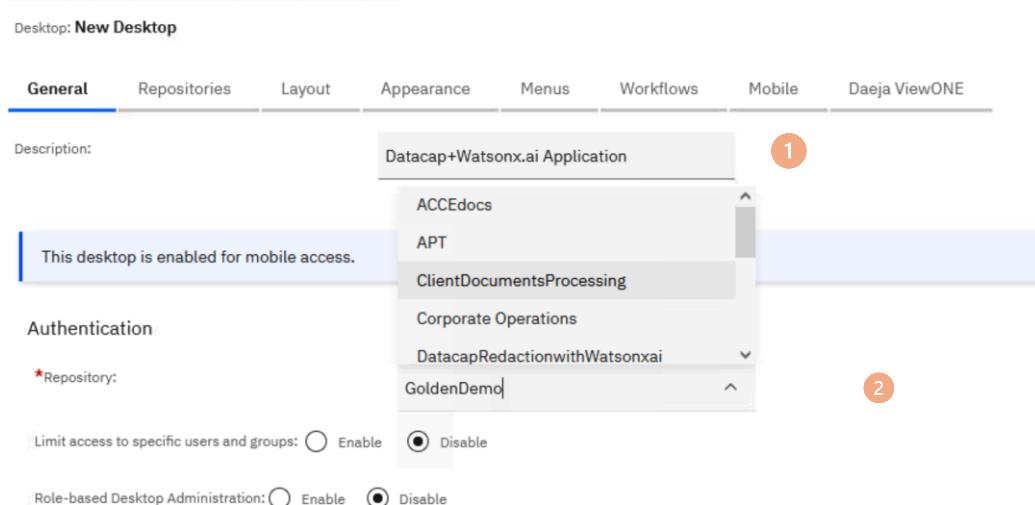
**New Desktop** **Edit** **Copy** **Delete** **Refresh** **Export** **Import** **Close**

Name	ID	Description
Admin Desktop	admin	Desktop for users with administrative privileges
Case Monitor	bawmonitor	IBM Business Automation Workflow
Datacap	datacap	Default desktop for Datacap Navigator
Datacap Admin Desktop	dcadmin	Datacap Navigator administration feature and Datacap Dashboard feature
Datacap Advanced Desktop	dcAll	Default desktop for Datacap Navigator, Datacap Navigator administration feature, and Datacap
2 Demo	dcgd	Datacap Golden Demo
Launch Desktop	dcQuickLaunch	Default desktop for Datacap Navigator with the quick launch pane
Lab	focusicnlab	Used for the ICN lab
New	focusnew	New ICN Layout
Focus ECM	focusecm	
IBM Administrative Console for Content Engine	acce	The desktop for IBM Administrative Console for Content Engine
IBM Business Automation Workflow	baw	Default desktop for IBM Business Automation Workflow Case Client

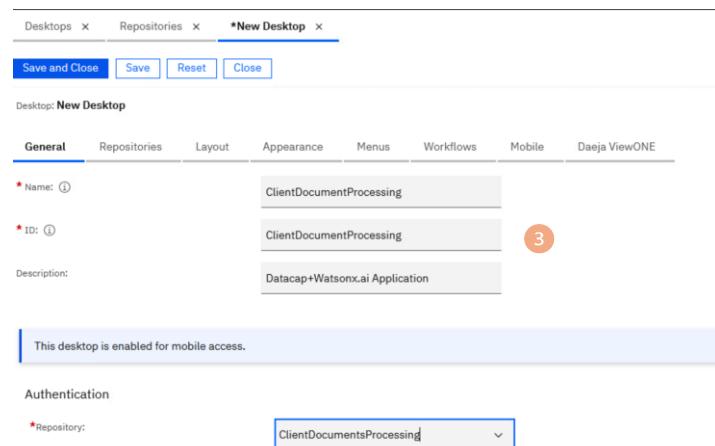
\_11. It opens the New Desktop option.

\_12. Select **ClientDocumentsProcessing** as the repository name for authentication.

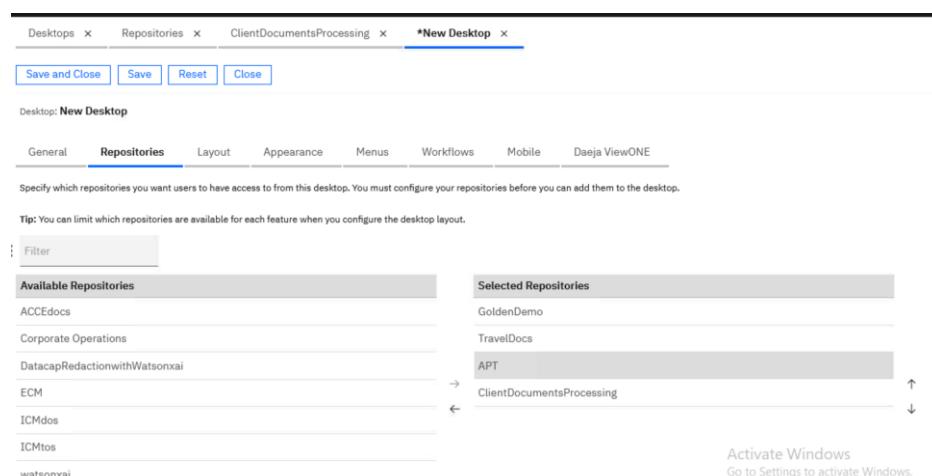
- **“Repository”:** From the drop down arrow select the ClientDocumentsProcessing repository created previously. This will be the default repository (application) when the desktop loads.



- **“Name”:** Add the Name you want to specify for your desktop. To make administration a bit easier, try and name it something like the repository name.
- **“ID”:** The ID you would like to call when loading the desktop in the URL.
- The rest can be kept as default.



\_13. Click on “**Repositories** Tab”. The Repository from the General tab will already appear in the **Selected Repositories**.



\_14. From the “Selected Repositories” list, remove all repositories except “ClientDocumentsProcessing Repo”—specifically, deselect **GoldenDemo**, **TravelDocs**, and **APT**. Then, click the arrow to move them to the “Available Repositories” section.

The screenshot shows the 'Repositories' tab selected in the top navigation bar. Below it, a tip states: "Tip: You can limit which repositories are available for each feature when you configure the desktop layout." A 'Filter' input field is present. On the left, a list of 'Available Repositories' includes: ACCEdocs, APT, Corporate Operations, DatacapRedactionwithWatsonxai, ECM, GoldenDemo, ICMdos, and ICMtos. On the right, a list of 'Selected Repositories' contains: ClientDocumentsProcessing. A blue-bordered arrow icon is positioned between the two lists, indicating the direction for moving items.

\_15. Go to the “Layout” tab.

The screenshot shows the 'Layout' tab selected in the top navigation bar. The 'Layout' section displays the current layout configuration: "ecm.widget.layout.NavigatorMainLayout". Under "Displayed features:", three items are listed: "Datacap Main Page", "Datacap Dashboard Page", and "Datacap Admin Console", all of which have checkboxes checked. To the right, a "Feature configuration" panel is visible with the message: "Select a feature to configure".

\_16. You will see that the “Display feature” options are already selected, inherited from the previously copied Datacap desktop.

\_17. Click **Save and Close**

You can create multiple desktops to give different users access to the content they need. For example, you can create a desktop that only allows users to access the ClientDocumentsProcessing application.

New Desktop Edit Copy Delete Refresh Export Import Close

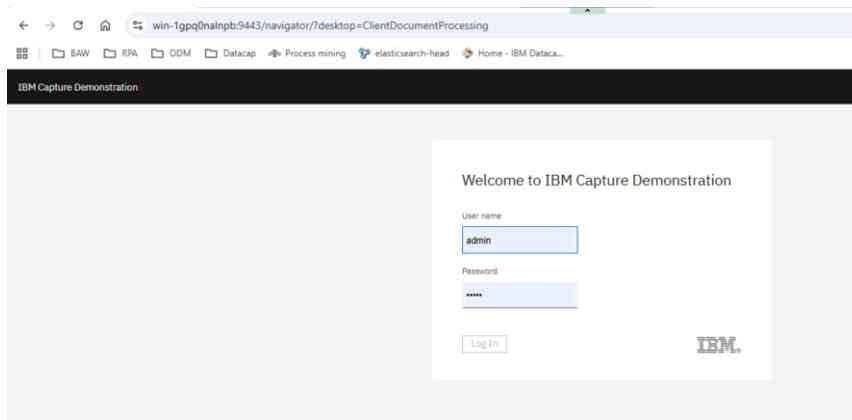
Name	ID
Admin Desktop	admin
Case Monitor	bawmonitor
ClientDocumentsProcessing	ClientDocumentsProcessing

>Login to the newly created Datacap desktop configured for the ClientDocumentsProcessing application.

\_18. Copy and paste the following URL into your web browser to access the **ClientDocumentsProcessing** desktop:

<https://win-1gpg0nalnpb:9443/navigator/?desktop=ClientDocumentProcessing>

Credentials are **admin/admin**



\_19. The Datacap view displays:

win-1gpg0nalnpb:9443/navigator/?desktop=ClientDocumentProcessing

IBM Capture Demonstration ClientDocumentsProcessing

Start Edit Job Edit Batch Delete View History

Queue	Batch	Job	Task	Status	Job Start	Job Time	Operator	Station	Docum...	Pages
3	20250916.000000	Demo_SingleTIFFs	Profiler	running	9/16/2025, 8:08 PM	0	admin	1	0	0
2	20250916.000001	Demo_SingleTIFFs	Export	aborted	9/15/2025, 9:09 PM	277	admin	1	0	0
1	20250916.000000	Demo_SingleTIFFs	Profiler	hold	9/15/2025, 8:37 PM	2383	admin	1	0	0

Filter Properties

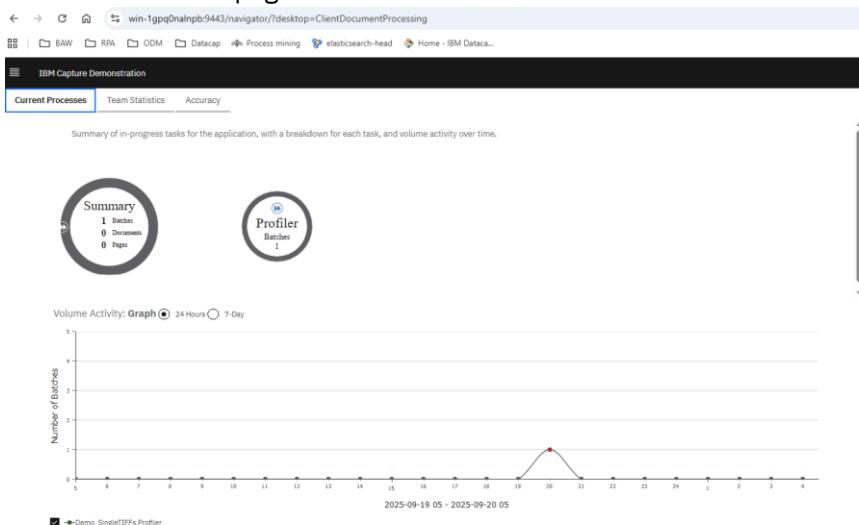
Class: Queue  
Batch: 20250916.000000

System Properties

Click on the **hamburger menu** to access the **Datacap Main Page** for business users and the **Datacap Admin Console** for administration. Since you're logged in as **admin/admin**, you have full privileges, which is why you can see both the **Admin Console** and the **Dashboard** page, where statistics are displayed.

Queue	Batch	Job	Task	Status	Job Start	Job Time	Operator	Station	Document	Pages
3	20250916.000000	Demo_SingleTIFFs	Profler	running	9/16/2025, 8:08 PM	0	admin	1	0	0
2	20250915.000001	Demo_SingleTIFFs	Export	aborted	9/15/2025, 9:09 PM	277	admin	1	0	0
1	20250918.000000	Demo_SingleTIFFs	Profler	hold	9/15/2025, 8:37 PM	2383	admin	1	0	0

## \_20. Check the Dashboard page to show the statistics



## \_21. Click on the down arrow to expand the Navigation Pane and select “Datacap Admin Console”

- On the opening Administration console, click “Users” on the left
- Click “**New User**”. Use **dadmin / dadmin** as Name and password:

IBM Content Navigator

Datacap Admin Console

Workflows

Groups

Users (1)

Stations

Shortcuts

Panels

Redaction Reasons

Workflows x Users x \*New User x (2)

Save and Close Save Reset Close

User: New User

General Permissions Privileges

\* Name (p8admin)

Description

\* Password

\* Retype Password

\_22. Switch to the “Permissions” and “Privileges” tab and allow this user to have all permissions and privileges:

The screenshot shows two side-by-side configuration panels. Both panels have tabs at the top: 'General', 'Permissions' (which is highlighted with a red box), and 'Privileges' (which is also highlighted with a red box). The left panel is titled 'Applications and Tasks' and lists various tasks with checkboxes. The right panel is titled 'Component' and lists various system components with checkboxes. Both panels include a 'Description' column for each item.

Applications and Tasks	Description
MGApp	Process Single Page TIFFs
Demo_SingleTIFFs	Process Single Page TIFFs
VScan	Run VScan Rules
Profiler	Recognize/Validate w/Rules
Verify	Verify with Rule Validation
Export	Export via Rules
Demo_WebScan	Import TIFFs from TMWeb
iVScan	Taskmaster Web VScan task
Upload	Upload images via Taskmaster Web
Profiler	Recognize/Validate w/Rules
Verify	Verify with Rule Validation
Export	Export via Rules
Demo_MultiFormat	Multi format images
VScanMulti	VScan for multiple formats

Component	Description
Job Monitor	
View Job Monitor	
More than one Job Monitor	
Delete batches	
Status change or rollback	
Priority or operator change	
Batch attributes change	
Change Job Attributes	
Save layout or set filter	
Administrator	
Workflow	
User groups	
Users	
Change passwords	

\_23. Click “Save and Close” and logout of Datacap Navigator.

### 3.3.3 Overview of new Datacap Desktop

\_1. Open Firefox and enter the url for the newly created desktop:

<https://win-1gpg0nalnpb:9443/navigator/?desktop=ClientDocumentProcessing>

\_2. Log in with User name: **dadmin** and password: **dadmin**

We are using an account with administration privileges. Your end users can be restricted using Datacap security features so that they only can access the capabilities that are authorized to use.

\_3. The Datacap view displays:

Review the screen and its primary sections:

This page is configurable and has four primary sections. For a particular desktop you can configure which panels will display.

- **Job Monitor** – A list of the batches that are being processed by Datacap.

The screenshot shows the IBM Content Navigator interface with the Datacap Main Page selected. On the left, there is a 'Quick Launch' panel containing icons for Scan, Verify, Upload, and Fixup. Below it is a 'Saved Filter' section. The main area is titled 'Job Monitor' and displays a table of jobs. The table has columns: Queue ID, Batch, Job, Task, Status, Job Start, Job Time, Operator, Pages, and Documents. Two rows are visible: one for a 'Demographic' job and another for a 'Navigator Job'. To the right of the table is a 'System Properties' panel showing details like Queue ID, Job Name, and Job Start Time. A red box highlights the 'Job Monitor' table.

Queue ID	Batch	Job	Task	Status	Job Start	Job Time	Operator	Pages	Documents
11	20250914.00000	Demographic	Import	instd	9/19/2025, 8:28 AM	2025	admin	1	0
10	20250914.00007	Navigator Job	Export	Job done	9/16/2025, 9:54 PM	1	admin	1	1

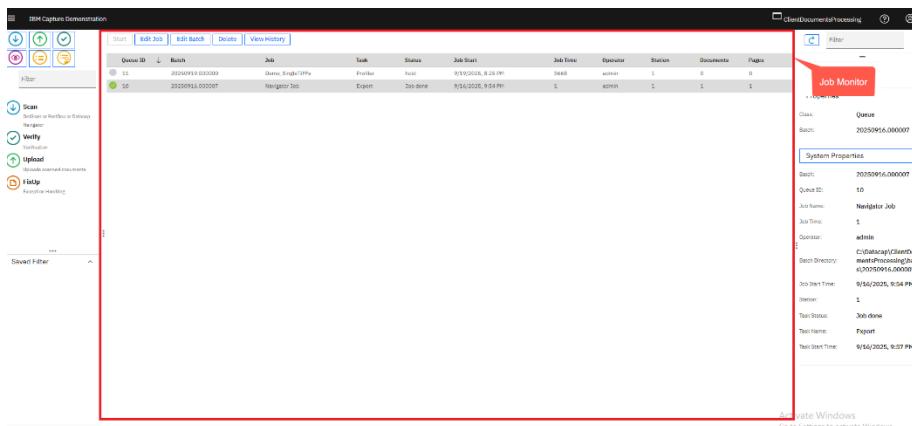
System Properties

Detail	Value
Queue ID	20250914.00007
Job Name	Navigator Job
Job Time	1
Operator	admin
Batch Directory	C:\Datacap\ClientDocuments\Processing\Batch\1\20250914.00007
Job Start Time	9/16/2025, 9:54 PM
Station	1
Task Status	Job done
Task Name	Export
Task Start Time	9/16/2025, 9:57 PM

- **Quick Launch Panel** – A single icon for each major type of task. This gives users a quick and easy list that launches tasks.

The screenshot shows the IBM Content Navigator interface with the Datacap Main Page selected. On the left, there is a 'Quick Launch' panel containing four icons: a blue circle with a downward arrow, a green circle with an upward arrow, a pink circle with an eye, and a yellow square with a colon and equals sign. A red box highlights this panel. To the right, there are buttons for 'Start' and 'Edit Job', and a 'Filter' input field. A red callout bubble points to the 'Quick Launch' panel with the text 'Quick Launch'.

- **Shortcut Panel** – A list of all the Datacap shortcuts. This gives the user a list of all the tasks that they are authorized to run.
- **Detail Panel** – Shows a thumbnail view and associated properties of the respective image when an item is selected in the Job Monitor.

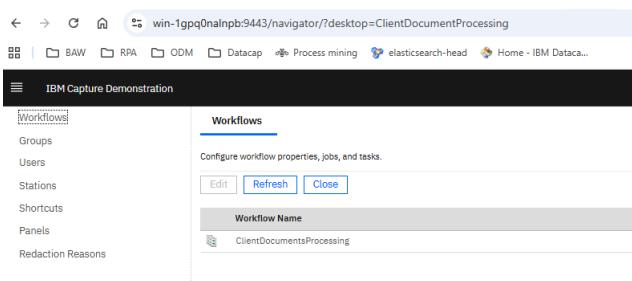


- Click on the first item in the Job Monitor list.
- Click on the blue triangle next to “System Properties”.

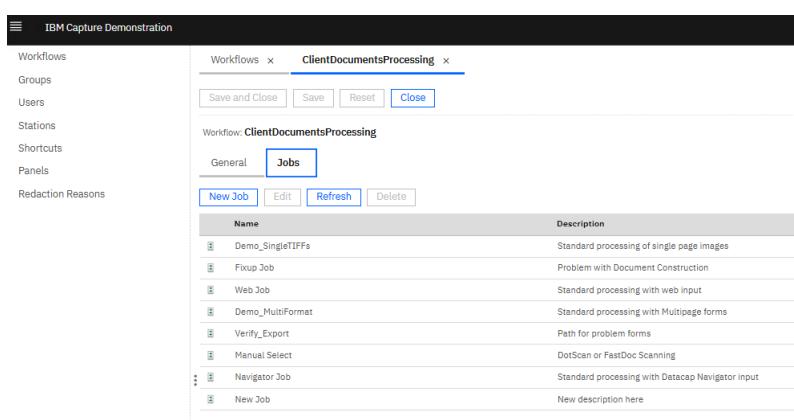
Note: You can expand and collapse the panels by clicking on the grey dots at the edge of the panels. You can also grab the edge of the panels and drag to resize the width.

If you skipped mapping the **verification ruleset** to the **Verify** task profile in step **1.10.2.3.2**, you can configure a new **Verify task** from the **Datacap Admin Console** within **Datacap Desktop** with following steps.

\_4. Click on **Workflows**, then select the **ClientDocumentsProcessing** workflow name from the list under **Workflows**.



\_5. Click on the Jobs tab.



\_6. Click on the **Navigator job** and select **Edit**, since this job will be used when processing batches as a **business user** in the Navigator interface.

Jobs	
<input type="button" value="New Job"/> <input type="button" value="Edit"/> <input type="button" value="Refresh"/> <input type="button" value="Delete"/>	
Name	Description
Demo_SingleTIFFs	Standard processing of single page images
Fixup Job	Problem with Document Construction
Web Job	Standard processing with web input
Demo_MultiFormat	Standard processing with Multipage forms
Verify_Export	Path for problem forms
Manual Select	DotScan or FastDoc Scanning
Navigator Job	Standard processing with Datacap Navigator Input

\_7. Click on the **New Task** button under the **Tasks** tab.

The screenshot shows the 'Navigator Job' workflow in the 'IBM Capture Demonstration' application. The 'Tasks' tab is active. A yellow box highlights the 'New Task' button in the toolbar below the task list. The task list itself is empty at this point.

\_8. Enter the following details as shown in the screen: Name the task **Verify** and select the program **verify.js**.

The screenshot shows the 'Verify' task configuration dialog. The 'General' tab is selected. The task details are as follows:

- Name: Verify
- Description: Verify with Rule Validation
- Mode: Normal
- Queue by: None
- Store: None
- Program: Verify.js

\_9. Click the Move Up button to position the Verify task before the Export task.

The screenshot shows the 'Demo\_MultiFormat' workflow tasks. The 'Tasks' tab is selected. The tasks listed are:

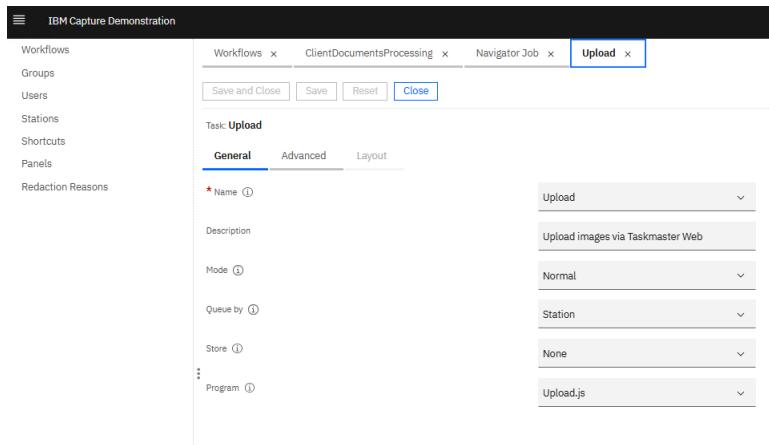
- VScanMulti
- PageID
- Profiler
- Verify
- Export

A yellow box highlights the 'Verify' task, indicating it has been moved up in the sequence.

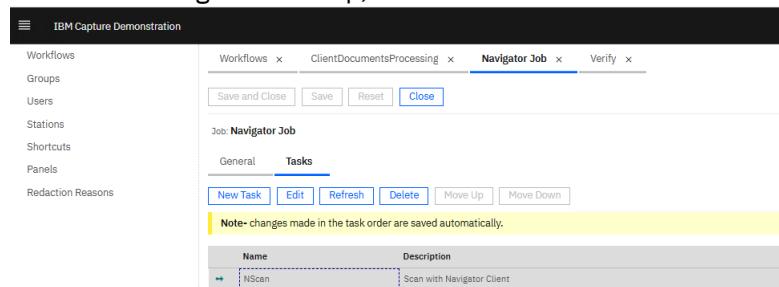
\_10. Click on the **New Task** button under the **Tasks** tab to add an **Upload Task** to the workflow. This task is required when scanning and processing sample documents from the **Datacap Navigator Desktop** in later sections.

\_11. Name the task **Upload** and select the program **upload.js**.

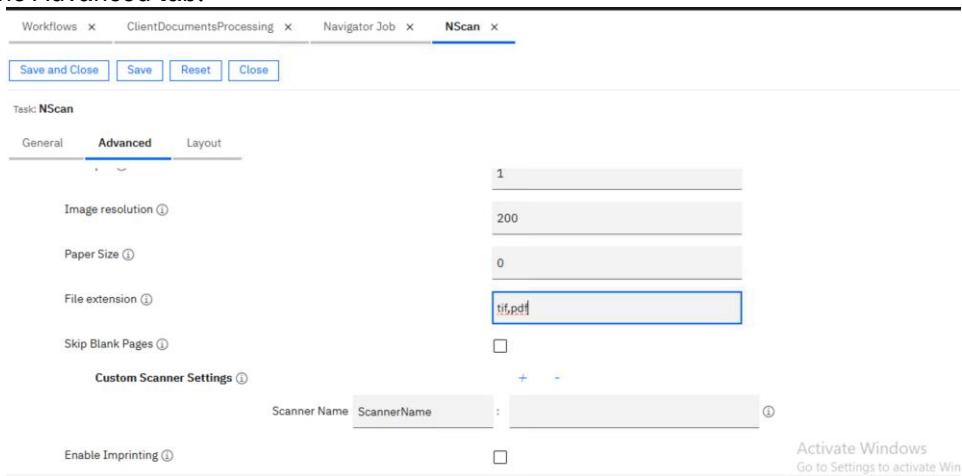
\_12. Click the Save and Close buttons once you have entered all the details, as shown below.



\_13. Go to the Navigator Job and click on the Nscan Task; since you will upload sample PDF documents to scan and process later from the Navigator desktop, be sure to add file extensions **tif** and **pdf**.



\_14. Click **Advanced**, search for **File Extension**, and ensure that you add the extensions **tif** and **pdf** in all relevant fields within the Advanced tab.



**15. Make sure your Navigator job workflow follows the order, mode, and program exactly as shown below.**

Name	Description	Mode	Program	Queue By
NScan	Scan with Navigator Client	Batch creation	Scan.js	None
Upload	Upload images via Taskmaster Web	Normal	Upload.js	Station
PageID	Page Identification Rules	Router	rulerunner.exe	None
Profiler	Recognize/Validate w/Rules	Router	rulerunner.exe	None
Verify	Verify with Rule Validation	Normal	Verify.js	None
Export	Export via Rules	Router	rulerunner.exe	None

\_16. Once you have added the file extensions, click **Save** and then **Close** to apply the changes.

### 3.3.4 RuleRunner Setup

In this section you are going to configure Rulerunner for background processing capability.

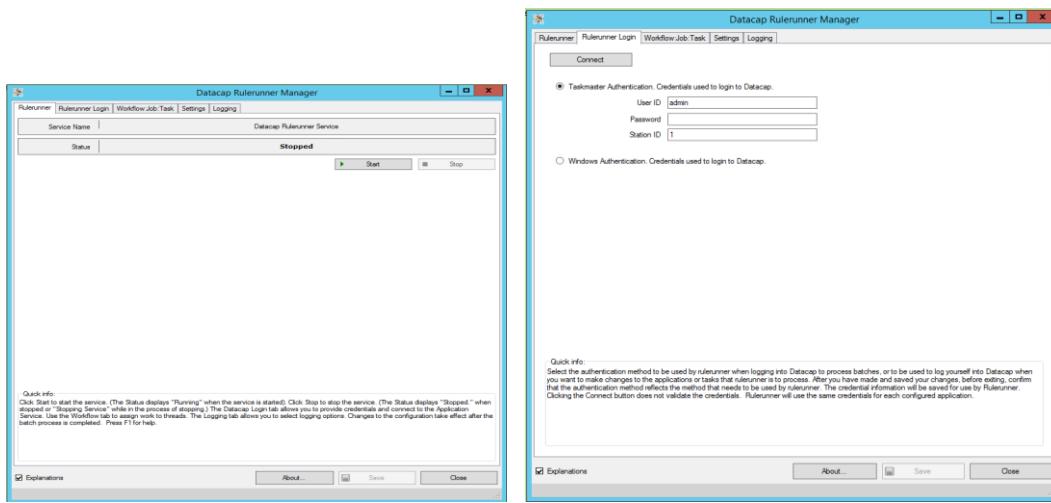
The Rulerunner service runs all tasks that do not require operator intervention, such as image cleaning, conversion, recognition, classification, and export to content repositories, such as FileNet Content Manager, IBM Content Manager, and other Content Management Interoperability Services (CMIS) compliant repositories.

Up to now you have manually processed tasks (Scan, Profile, Export) for your batches. With the Rulerunner service configured, tasks that do not require operator intervention can be processed in the background.

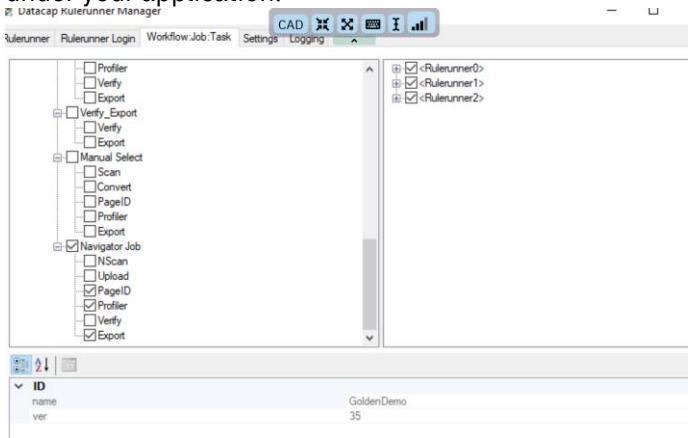
Rulerunner is normally run on a separate server for performance reasons. Multiple instances can be run on multiple servers and each instance can be configured to run multi-threaded for a multi-core processor. This helps complete the document processing faster and simultaneous processing results also in better resource usage.

#### Start RuleRunner Configuration – Step by Step

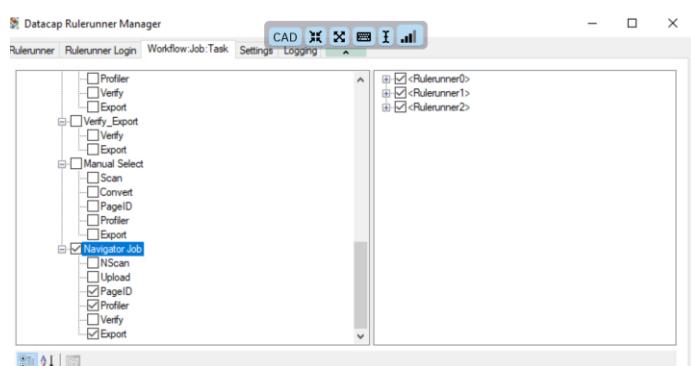
1. From the Workstation start menu select **Datacap Rulerunner Manager**.
2. The Rulerunner Manager interface comes up. If your status is Running, click the stop button.
3. Click on the “Rulerunner Login” tab.
4. At the login screen enter admin for user id and admin for password. Station ID should be “1”.



5. Click on Connect and select the “Workflow Job: Task” tab.
6. On left side are all the applications shown that are configured for Datacap. Click in the check box to the left of your application (CilentDocumentsProcessing). The tree will expand showing all the jobs created under your application.



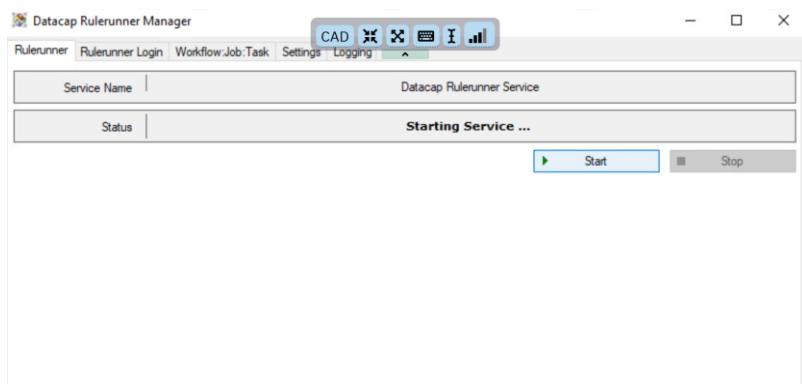
7. Select the **Navigator Job** and check only the following options to run as background tasks: **PageId**, **Profiler**, and **Export**.  
Then, drag and drop the **Navigator Job** onto the right-hand side, placing it on one of the **RuleRunner** threads.



8. You will see the selected tasks running in the background after you scan and upload a document from the **Navigator Desktop**.



9. Click **Save**, then start the **Rule Runner** server.



### 3.3.5 Scanning Documents

One of the primary functions of Datacap is scanning paper documents. Content Navigator supports the operation of scanners from the web browser. Scanners that use an industry TWAIN driver can be operated.

For this exercise, we don't have a scanner attached. Therefore, in this section we will use the import feature to simulate scanning. Datacap lets you select pre-scanned images from your computer's files in a manner similar to scanning documents.

#### Start Scan Task – Step by Step Instructions

- 1) Open Firefox and enter your Datacap Navigator desktop URL again
- 2) Login as **dadmin** with password **dadmin**.
- 3) Click on the **Scan** icon in the Quick Launch menu.



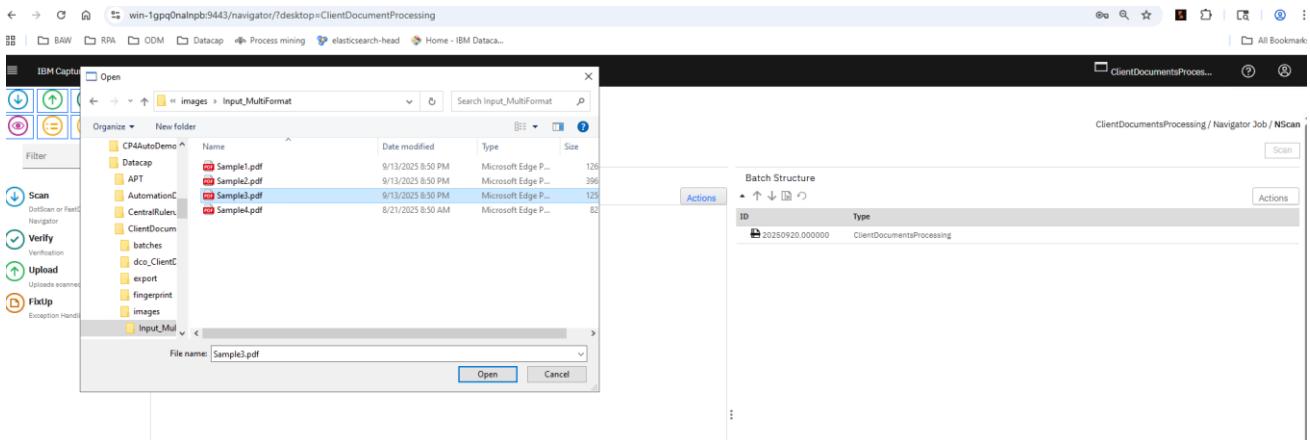
The scan task screen displays.

The screenshot shows the Datacap Main Page in the IBM Content Navigator. The interface is organized into three main sections:

- Input:** This section is highlighted with a red box. It contains a dropdown menu set to "Import from Directory" with a "Browse..." button and a "Scan" button.
- Scanned:** This section is highlighted with a purple box. It displays "Scanned Pages (0/0)" and a message "No images to display".
- Batch:** This section is highlighted with a yellow box. It shows the "Batch Structure" with one item: ID 20180628.000000 and Type MGAApp.

There are 3 sections:

- 1) Input Panel: Location of documents to be specified
  - 2) Scanned Pages: Image Viewer
  - 3) Batch Structure: The list of scanned pages
- 
- 4) Click on the **Browse** button. (Note if necessary navigate to the folder “C:\Datacap\ClientDocumentsProcessing\images\Input\_MultiFormat ”)
- 
- 5) Click on **Sample3.pdf** on the file list and then click on the **Open** button



This returns to the Scan Task screen which shows the selected folder name and file next to the Browse button.

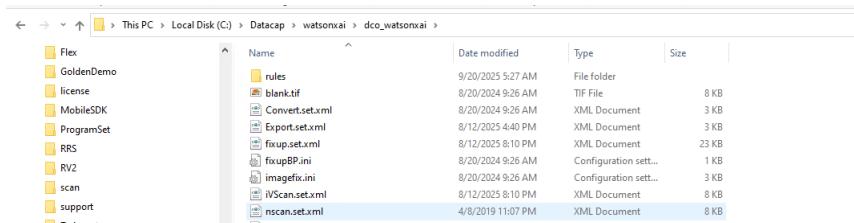


Click on the **Scan** button.

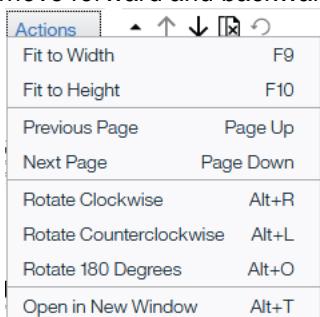
The image will be uploaded into the system and displayed on the screen. The “Batch Structure” shows a list of the pages that were scanned. The “Image Viewer” will always display the last image.

#### Troubleshooting:

It might be, that the buttons “Scan” and “Submit” are greyed out or that the preview of the scanned document is not shown. Copy the nscan.set.xml file from working datacap DCO folder to your Datacap application DCO folder



You can zoom in and zoom out by clicking on the magnifying glass buttons  
 From the “Actions” button of the image viewer you can fit the image into your frame, manipulate the image or move forward and backward in the batch.



6) Click on the **Submit** and click **upload icon** from left menu.



This saves the batch and you are returned to the job monitor.

We have configured this desktop to automatically upload the images to the Datacap server. Note: This automatic upload can be turned off from the navigator desktop.

**Hint:**

If the button is greyed out, please refer to “Troubleshooting” of step 6”.

7) Notice that your batch is in the list. The batch is at the “Profile” task and is pending (your batch number and id may be different).

The screenshot shows the Datacap Rulerunner interface. At the top, there are several icons for actions like Scan, Verify, and Upload. Below the toolbar is a table with columns: Queue ID, Batch, Job, Task, Status, Job Start, Job Time, Operator, Station, Documents, and Pages. A single row is selected, showing Queue ID 10, Batch 20250916.000007, Job Navigator Job, Task pending, Status pending, Job Start 9/16/2025, 9:54 PM, Job Time 0, Operator 0, Station 0, Documents 0, and Pages 1. To the right of the table is a 'Properties' panel showing Class: Queue and Batch: 20250916.

We have set up Datacap Rulerunner in the previous exercise to automatically process this task in the background.

- 8) Click on the refresh button to see update to the job monitor.

The server processes the batch: runs OCR, validates data, and creates separate documents; each containing one or more pages.

Notice the status changes from pending to running as the background processes execute.

When the batch is ready, the Task to perform will change to “Verify” and the status will change to pending.

This screenshot shows the IBM Capture Demonstration interface, specifically the ClientDocumentProcessing tab. It has a similar layout to the previous screenshot, with a toolbar, a table of jobs, and a properties panel. The table shows a single row with Queue ID 32, Batch 20250920.000000, Job Navigator Job, Task Verify, Status pending, and other details. The properties panel on the right shows the same information as the previous screenshot.

### 3.3.6 Verify a Batch

At this point Datacap has processed the scanned images. It did the following steps in background:

- Enhanced the images
- Identified the types of pages
- Assembled the pages into documents
- Read data from the images using Optical Character Recognition (OCR)
- Flagged images and data that require human review (Verify)

This exercise walks through the human review or “Verify” task.

Since we are reading data from paper documents the results may not be 100% accurate. For example, pages may be damaged, users can enter invalid information on paper or simply cross out or erase information on the paper.

As a result, the system checks the data and flags fields that need human review. To be conservative, the system may flag some fields that are actually OK. When you configure the system, you can set the threshold for the acceptable level of confidence.

Fields that are low confidence are displayed in yellow. Fields that have errors are displayed in red.

The operator reviews each page to correct any problems. If a page has no problems the system can be configured to skip it, so that the user only reviews problems.

Start the Verify task – Step by Step instructions

- 1) Double click on your batch that is pending for “Verify”.

The scan task screen displays. There are 3 sections:

- Image Viewer: Image Viewer
- Field Details: Data entry panel
- Batch Structure: The list of scanned document and pages

Observe that the system has lifted data from the scanned images. You may also see that some fields are highlighted yellow. This indicates that the system is not fully confident that the data is correct, therefore it is marked for you to verify. You can move to the next problem or entry with low confidence within a batch via the buttons at the top.

[Submit](#) [Hold](#) [Previous Page](#) [Next Page](#) [Previous Problem](#) [Next Problem](#) [Next Low Confidence](#) [Run Validations](#)

Above each field is a “snippet”. A snippet is a small cutout area of the image where the data is located. This lets you see the image data easily without looking at the image display.

On the right is the list of documents and the pages inside of the documents. The system determined where each document started and what pages were contained in each document. Documents and pages are marked as “OK” or “Problem”. The operator is only required to look at the problem areas. If your data in the fields are correct (no low confidence nor validation error) it does not need to be changed.

2) Click on **Submit** once you have verified the image.

[Submit](#) [Hold](#) [Previous Page](#) [Next Page](#) [Previous Problem](#) [Next Problem](#) [Next Low Confidence](#) [Run Validations](#)

3) Click OK to finish the batch

Submit

All documents are complete. Do you want to finish batch?

[OK](#) [Cancel](#)

When the Job Monitor redisplays, the batch will now show that it is pending for the “Export” task.

The screenshot shows the IBM Capture Demonstration interface. In the center, there is a table with columns: Queue ID, Batch, Job, Task, Status, Job Start, Job Time, Operator, Station, Documents, and Pages. A single row is selected, showing Queue ID 12, Batch 20250920.000000, Job Navigator Job, Task Export, Status pending, Job Start 9/20/2025, 8:06 AM, Job Time 0, Operator admin, Station 1, Documents 1, and Pages 1. Above the table, there are buttons for Start, Edit Job, Edit Batch, Delete, and View History. To the left, there is a toolbar with icons for Scan, Copy, Paste, and others. To the right, there are tabs for Properties and Queue, and a sidebar with a 'Scan' section.

The “Export” task outputs the data to a Datacap database and stores the Document in FileNet.

When the Export is completed, the status will change to Job Done.

This screenshot is identical to the previous one, but the status of the selected job has changed from "pending" to "Job done". The rest of the interface remains the same, including the table data and the sidebar.

Note: You have to refresh your monitor to see Job Done.

### 3.3.7 Viewing Results in Repository

\_1. Log in to the **Focus Corp** application using the bookmarked favorite under **BAW → Desktops**.

The screenshot shows the BAW desktops menu. Under the "Desktops" heading, there is a list of applications: Workflow Center - Solutions, Process Center, Case Builder, ACCE Console, Process Admin Console, Process Inspector, Workplace, Process Portal, REST API Tester, Swagger UI, WebSphere Integrated Solutions Console, and Swagger UI OPS. A dropdown menu is open over the "Focus Corp" item, showing its URL: https://win-1gpq0nalnpb:9443/navigator/?desktop=focusecm. To the right of the dropdown, there is a "Welcome" panel with fields for "User name:" (dadmin) and "Password:", and a "Login" button.

\_2. Click on the ClientDocuments from the left hand menu and you can see folders list

The screenshot shows the Focus Corp interface. On the left, there is a navigation tree with nodes: ECM, Case Folders, ClientDocuments (which is expanded), Departments, and others. The main area shows a table with a single column "Name" containing five entries: 20250918.000005, 20250918.000007, 20250919.000000, 20250919.000000, and 20250920.000000. At the top, there are buttons for Add Document, Refresh, Add Document, New Folder, New, and Actions.

\_3. Click on **Batch**, then open the exported document as shown below.

The screenshot shows the Focus Corp ECM interface. On the left, there's a navigation tree with categories like ECM, Case Folders, and ClientDocuments. In the center, a list of documents is displayed, with one document selected: 'tm000001'. On the right, a detailed view of the selected document is shown, including its properties:

Class:	LegacyBankingInformation
Document Title:	tm000001
ClientName:	Legacy Consulting
BankPhoneNumbe:	+1-213-111-7890
BankName:	XYZ Bank

\_4. You can view the document properties that were extracted using the **watsonx.ai LLM model**.

The screenshot shows the Focus Corp ECM interface with the same document selection. A new window titled 'Focus Corp Viewer - Google Chrome' is open, showing the extracted banking information from the document:

Banking Information

This banking information provided here will be used by Focus Corp to withdraw the monthly services fee.

Client Name	Legacy Consulting
Bank Name	XYZ Bank
Bank Phone Number	+1-213-111-7890
Bank Address	887 Cypress Rd, Garden Grove, CA 92840

The properties panel on the right remains the same as in the previous screenshot.

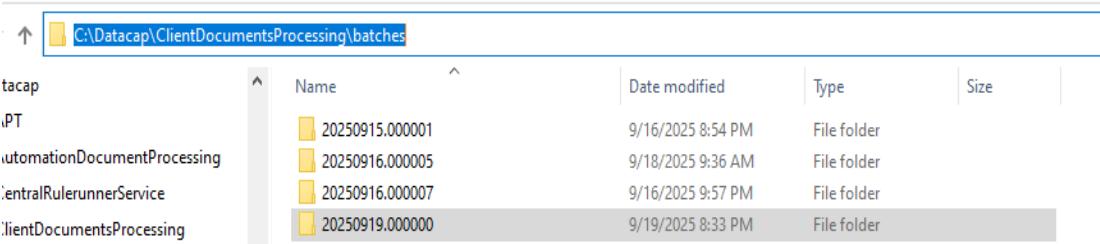
\_5. Log out and close the browser.

## 4. Troubleshooting

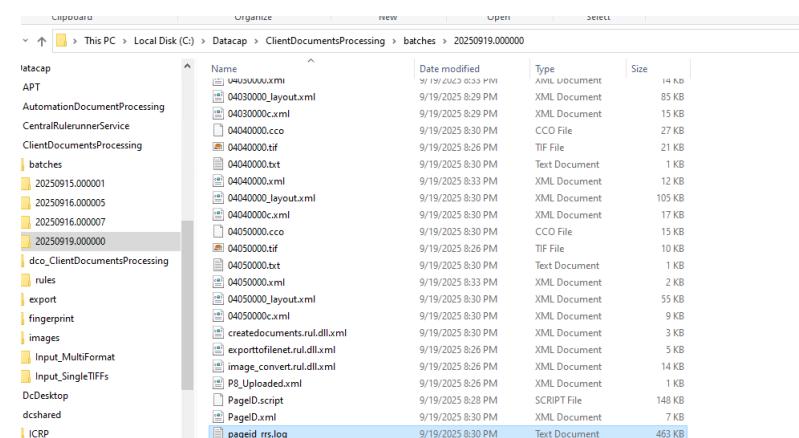
In this section, you will explore how to **troubleshoot issues** related to **classification and extraction failures** when using **Watsonx.ai**, including verifying whether the prompt was properly sent to the model. You can also **validate the prompt and its response** using the **Watsonx.ai Free Form Style tile** for further analysis.

If you get stuck or encounter with an issue at the **classification stage**, navigate to the batch folder at C:\Datacap\ClientDocumentsProcessing\batch, then open the specific **problem batch folder**, which is named using the format **YYYYMMDD.HHMMSS**.

\_1. For example, if the batch **20250919.00000** encountered a problem at the **classification stage**, open that specific batch folder to begin troubleshooting.

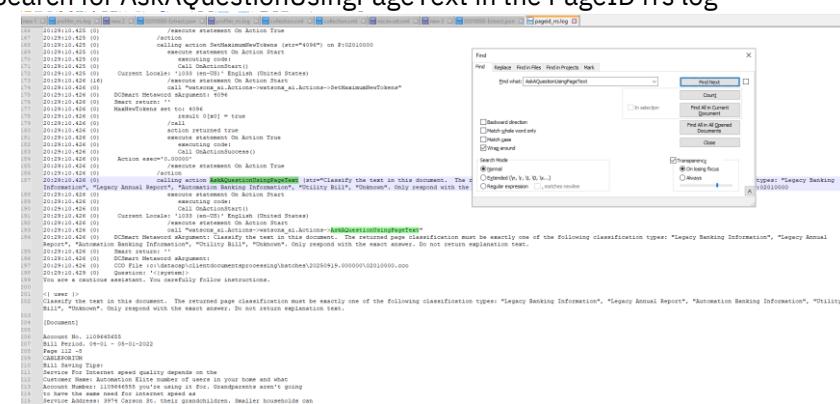


\_2. Open the **PageID\_rrs.log** file in notepad++ which is located within the batch folder, where the **recognition** and **Watsonx.ai PageID ruleset** were defined, to review the classification process and identify any issues.



When **enhanced logging** is enabled in the watsonx.ai\_PageID ruleset, you can view the **prompt sent** and the **response received** from the **Watsonx.ai SaaS platform**, allowing you to verify the interaction details for troubleshooting or validation.

\_3. Search for AskAQuestionUsingPageText in the PageID rrs log



\_4. Check the **prompt** that was sent and the **response** received from **Watsonx.ai** by reviewing the log file (e.g., **PageID\_rrs.log**), to ensure the request was correctly formatted and the model returned the expected output.

```
20250919.0000 (0)    CCU File :c:\datcap\clientdocumentsprocessing\batches\20250919.000000\02010000.ccc  
20250919.049 (0)    Question: <system>  
ou are a cautious assistant. You carefully follow instructions.  
  
[ user ]>  
classify the text in this document. The returned page classification must be exactly one of the following classification types: "Legacy Banking Information", "Legacy Annual Report", "Automation Banking Information", "Utility Bill", "Unknown". Only respond with the exact answer. Do not return explanation text.  
  
[ Document]  
  
account No. 1109646565  
bill Period: 04-01 - 05-01-2022  
age 122 -8  
amount 62.76  
bill Savings Tip:  
service For Internet speed quality depends on the  
customer Name: Automation Elite number of users in your home and what  
account Number: 1109646555 you're using it for. Grandparents aren't going  
to have the same kind of internet speed as  
service address: 3574 Carson St. their grandchildren. Smaller households can  
enjoy, MI 48911 have faster web browsing, music, and HD.  
bill Bill Period Due On 04-01-2022 05-05-2022 - 05-01-2022 streaming gaming, download and by speed. HD using video Email, as little streaming web as 18 browsing, for Mbps three of  
ervices can be achieved with as little as 30  
Mbps of download speed.  
method: POST  
headers: Content-Type: application/json  
Body: {  
  "prompt": "What are the different ways to pay?  
  previous Balance $ 60.05 You may pay your bill by sending a check  
  using the enclosed return envelope. You may  
  payment Received - $ 60 .05 also opt to pay with your credit or debit card  
  new Charges - see details $ 62.76 via our website. You can also enroll your  
  local Account Due $ 62.76 card convenience. For auto debit option for your  
  local account.  
  bill Period 04-01-2022 - 05-01-2022  
  bill Due On 05-05-2022  
  AMLEPORIUM Amount Due. $ 62.76  
  algorithm  
  0 BOX  
  0 Box 1 Account Enclosed  
  mpano Beach, Florida 33060 US  
  hill Automation Elite  
  974 Carson St.  
  ensing, MI 48911  
}  
  
End1
```

```
20250919.059 (0)    Bearer response was deserialized  
20250919.059 (0)    refresh not supported, type: Bearer, expiration: 1758342547  
20250919.059 (0)    Obtained api token  
20250919.060 (14)    Bearerkey factory  
20250919.060 (0)    Bearer build post  
20250919.060 (406)    token received  
20250919.076 (0)    API-Complete-Token: 416  
20250919.076 (0)    Bearer response was deserialized  
20250919.076 (0)    refresh not supported, type: Bearer, expiration: 1758342548  
20250919.076 (0)    Building Client  
20250919.076 (0)    Building Client  
20250919.076 (0)    Model: "mistrailai/mistrail-small-3-1-24b-instruct-2503", URL: "https://us-south.ml.cloud.ibm.com/ml/v1/text/generation?version=2023-05-29", decode: 'greedy', Temp: 1.0, mTokens: 4096, TimeOut: 0, PID: 1023b96e-9139-427d-bfde-37e4141bdbe200  
20250919.076 (0)    Added bearer and ID  
20250919.076 (0)    Starting plugin request  
20250919.246 (198)    Response: Status: 1:1, Content: System.Net.Http.StreamContent, Headers:  
{  
  Connection: keep-alive  
  Server-Timing: on�id=ticks=1e207b2700f76ca  
  Pragma: no-cache  
  X-Request-Limit-Rate: 8  
  X-Request-Id: e645ff0f022cc5b2d825a7a6b6clf16c  
  X-Xss-Protection: 1; mode=block  
  X-Xss-Protection: 1; mode=block  
  X-Request-Limit-Period: 1000  
  X-Request-Limit-Reset: 7  
  X-Request-Limit-Reset-After: 125  
  Strict-Transport-Security: max-age=31536000; includeSubDomains  
  Strict-Transport-Security: max-age=31536000; includeSubDomains  
  X-Frame-Options: DENY  
  Referrer-Policy: strict-origin  
  X-Content-Type-Options: nosniff  
  Content-Security-Policy: default-src 'none'; script-src 'self'; connect-src 'self'; img-src 'self'; style-src 'self'; frame-ancestors 'none'; form-action 'self';  
  of-cause-status: DYNAMIC  
  Cf-Ray: 5b1e3a3dd5d9d4-IAD  
  Cache-Control: no-store, no-cache, must-revalidate, no-cache  
  Date: Sat, 20 Sep 2025 03:25:11 GMT  
  Server: cloudflare  
  Content-Length: 1136  
  Content-Type: application/json  
}  
20250919.265 (0)    Response Body: {"model_id": "mistrailai/mistrail-small-3-1-24b-instruct-2503", "model_version": "1.0.0", "created_at": "2025-09-20T03:29:11.222Z", "results": [{"generated_text": "Utility Bill", "generated_token_count": 3, "input_token_count": 615, "stop_reason": "eos_token"}]} *warning*: This model is a Non-IBM Product governed by a third-party license that may impose use restrictions and other obligations. By using this model you agree to its terms as identified in the following URL: *.id*: "disclaimer_warning", "more_info": "https://dataplatform.cloud.ibm.com/doc/content/rex_analyze-data/fm-models.html?context=ryw", "message": "In future implementation, the parameter 'parameters.decoding_method' will be ignored and set automatically." id*: "param_deprecation", "message": "The value of 'parameters.time_limit' for this model must be larger than 0 and not larger than 10mS; it was set to 1ms. Instead, use 'parameters.time_range', 'additional_properties':{"limit":600000, "new_value":600000, "parameter": "parameters.time_limit", "value":0}, (*message*: "This API is legacy. Please consider using '/ml/v1/text/chat' instead.", "id": "api legacy")}  
20250919.266 (0)    API-Complete: 836  
20250919.266 (0)    Full results:  
[  
  {"generated_text": "Utility Bill",  
   "generated_token_count": 3,  
   "input_token_count": 615,  
   "stop_reason": "eos_token"}]
```

Activate Windows  
Go to Settings to activate Windows.

\_5. Sometimes, you may encounter errors if the API Key, Project ID, or Model Name are not passed in the correct order during the call, or if you're out of tokens, or if the selected model is not available in the region where your Project ID was created.

\_6. For example, if the batch **20250919.0000** encountered a problem at the **extraction stage**, open that specific batch folder to begin troubleshooting, then open the **Profiler\_rrs.log** file and search for **errors**, as well as the **prompt request** and **response** from Watsonx.ai.

If your prompt does not meet your requirements, you can **tune or refine it** directly on the **Watsonx.ai platform**, using the **Free Form Prompting** interface as shown below.

\_7. You can tune your prompt by copying it from the **PageID\_rrs.log** or **Profiler\_rrs.log** files and refining it directly on the Watsonx.ai platform.

\_8. Copied the classification prompt from the **PageID\_rrs log** and submitted it to generate a response from the LLM model for verification.

< user >  
Classify the text in this document. The returned page classification must be exactly one of the following classification types: "Legacy Banking Information", "Legacy Annual Report", "Automation Banking Information", "Utility Bill", "Unknown". Only respond with the exact answer. Do not return explanation text.

[Document]

Account No. 1109645655  
Bill Period. 04-01 - 05-01-2022  
Page 112 -8  
CABLEPORUM  
Bill Saving Tips:  
Service For Internet speed quality depends on the Customer Name: Automation Elite number of users in your home and what Account Number: 1109645655 you're using it for. Grandparents aren't going to have the same need for internet speed as Service Address: 3974 Carson St, their grandchildren. Smaller households can Lansing, MI 48911 have email web browsing, music, and HD Bill Bill Period Due On 04-01-2022-05-05-2022 - 05-01-2022 streaming gaming, download and by speed. HD using video Email, as little streaming web as 18 browsing, for Mbps three of devices can be achieved with as little as 30 Mbps of download speed.  
Monthly Statement Summary Ways to Pay:  
Previous Balance \$ 60.05 You may pay your bill by sending a check using the enclosed return envelope. You may Payment Received - \$ 60.05 also opt to pay with your credit or debit card New Charges - see details \$ 62.76 via our website. You can also enroll your Total Amount Due \$ 62.76 card convenience, for auto debit option for your 8 Account Number 1109645655  
Bill Period 04-01-2022 - 05-01-2022  
Bill Due On 05-05-2022  
CABLEPORUM Amount Due. \$ 62.76  
Call Us  
PO BOX 1108 Amount Enclosed  
Pompano Beach, Florida 33060 US  
Ihill Automation Elite

Clear output Generate

< user >  
Page 112 -8  
CABLEPORUM  
Bill Saving Tips:  
Service For Internet speed quality depends on the Customer Name: Automation Elite number of users in your home and what Account Number: 1109645655 you're using it for. Grandparents aren't going to have the same need for internet speed as Service Address: 3974 Carson St, their grandchildren. Smaller households can Lansing, MI 48911 have email web browsing, music, and HD Bill Bill Period Due On 04-01-2022-05-05-2022 - 05-01-2022 streaming gaming, download and by speed. HD using video Email, as little streaming web as 18 browsing, for Mbps three of devices can be achieved with as little as 30 Mbps of download speed.  
Monthly Statement Summary Ways to Pay:  
Previous Balance \$ 60.05 You may pay your bill by sending a check using the enclosed return envelope. You may Payment Received - \$ 60.05 also opt to pay with your credit or debit card New Charges - see details \$ 62.76 via our website. You can also enroll your Total Amount Due \$ 62.76 card convenience, for auto debit option for your 8 Account Number 1109645655  
Bill Period 04-01-2022 - 05-01-2022  
Bill Due On 05-05-2022  
CABLEPORUM Amount Due. \$ 62.76  
Call Us  
PO BOX 1108 Amount Enclosed  
Pompano Beach, Florida 33060 US  
Ihill Automation Elite  
3974 Carson St.  
Lansing, MI 48911

[End]  
Utility Bill

Stop reason: End of sequence token encountered  
Tokens: 590 Input + 4 generated = 594 out of 128000  
Time: 0.3 seconds

Clear output Generate

\_9. Copied the Extraction prompt from the profiler\_rrs log and submitted it to generate a response from the LLM model for verification.

< user >  
Account Number  
7250512345  
Account Type  
Ø Checking × Savings Ø Other  
By submitting this form, you authorize Focus Corp to withdraw the monthly services fee from the specified bank account.  
Date  
5/10/21

[End]  
The following is a list of keys to find within the document. Please find only these keys and do not return any additional keys.

Client Name  
Bank Name  
Bank Phone Number  
Bank Address  
Routing Number  
Account Number  
Account Type  
Date

Return the key value pairs in JSON format. Return only JSON in the following format. Do not include additional keys. Do not include additional text before or after the JSON.

```
{<assistant>:  
  "Client Name": "Legacy Consulting",  
  "Bank Name": "XYZ Bank",  
  "Bank Phone Number": "+1-213-111-7890",  
  "Bank Address": "887 Cypress Rd, Garden Grove, CA 92840",  
  "Routing Number": "1234567890123456",  
  "Account Number": "7250512345",  
  "Account Type": "Checking",  
  "Date": "5/10/21"}  
}
```

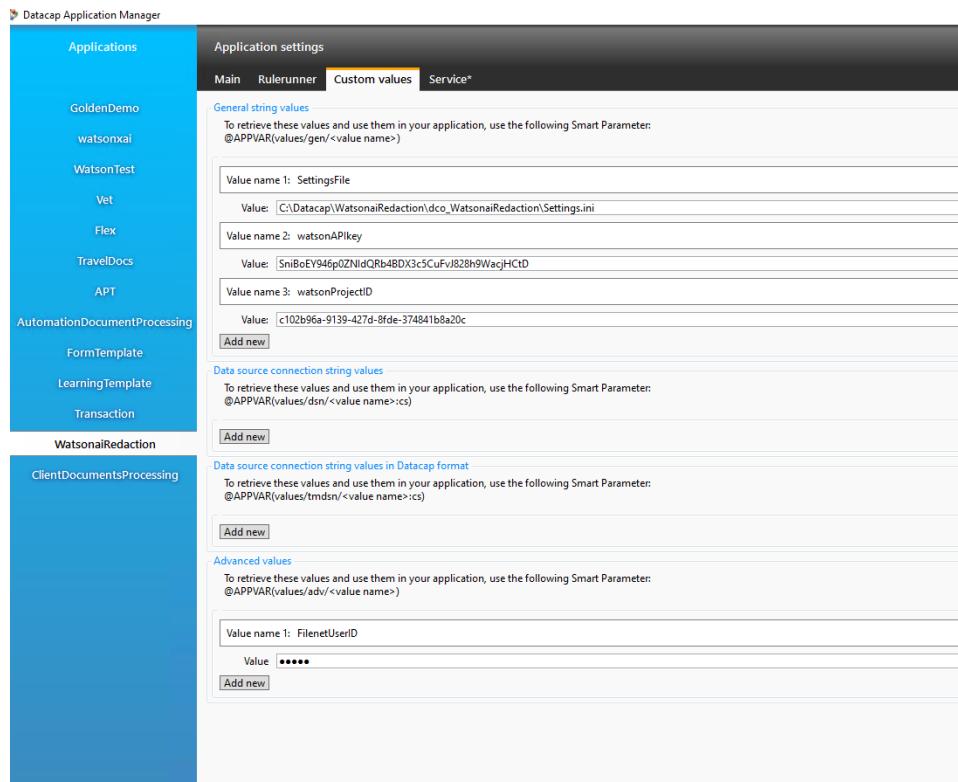
Stop reason: End of sequence token encountered  
Tokens: 314 Input + 122 generated = 436 out of 128000  
Time: 3.5 seconds

Clear output Generate

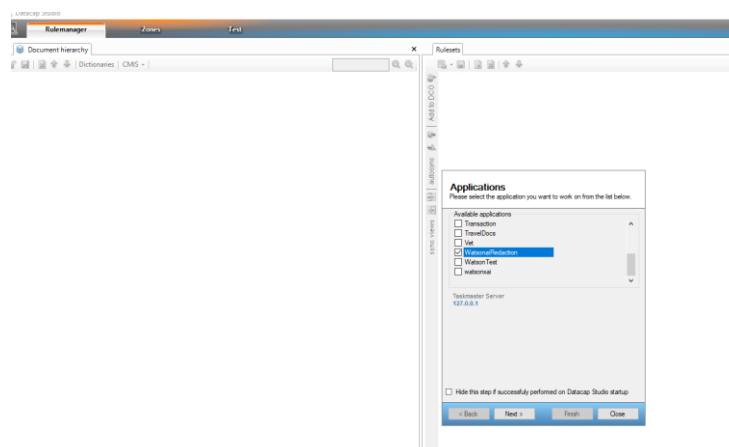
## 5. Explore Document Redaction with Watsonx.ai

The \*\*WatsonxaiRedaction\*\* application has been prepared to redact sensitive data from medical records, demonstrating how data redaction can be performed using the Watsonx.ai LLM model. You can open this application in \*\*Datacap Studio\*\* to explore the rulesets, rules, and functions implemented, see how the rules are mapped to the DCO, and test the application as part of this lab exercise for your hands-on awareness.

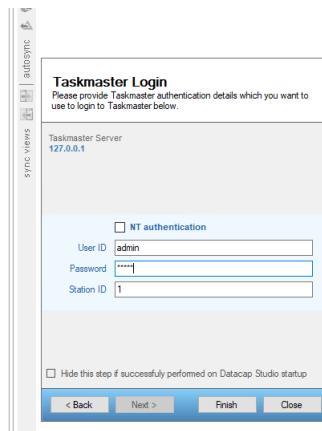
\_1. Before testing the WatsonxaiRedaction app in Datacap Studio, you need to set the **ProjectID** and **APIKey and save changes** in the Datacap Application Manager under **watsonxaiRedaction** within the **Custom Values** section, as shown below.



\_2. Login to watsonxaiRedaction app from the datacap studio.



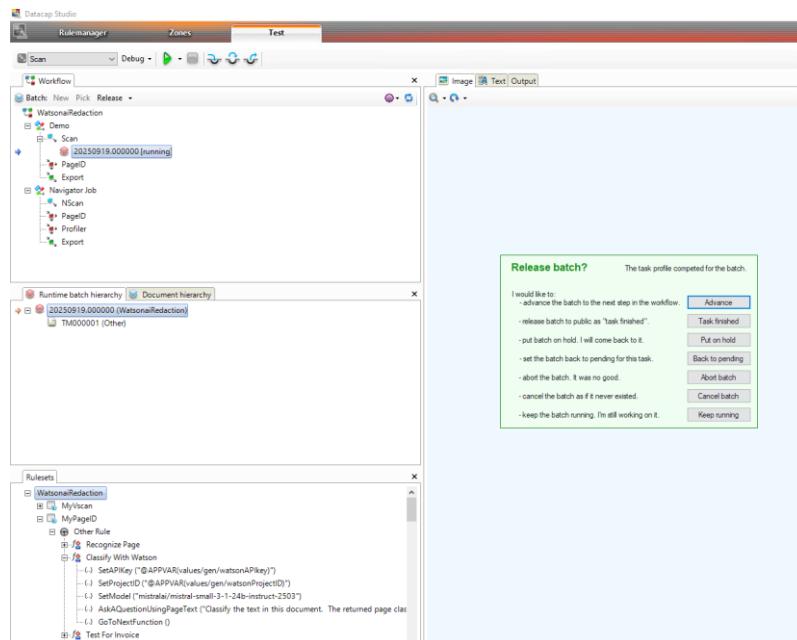
\_3. Provide login credentials admin/admin and click finish



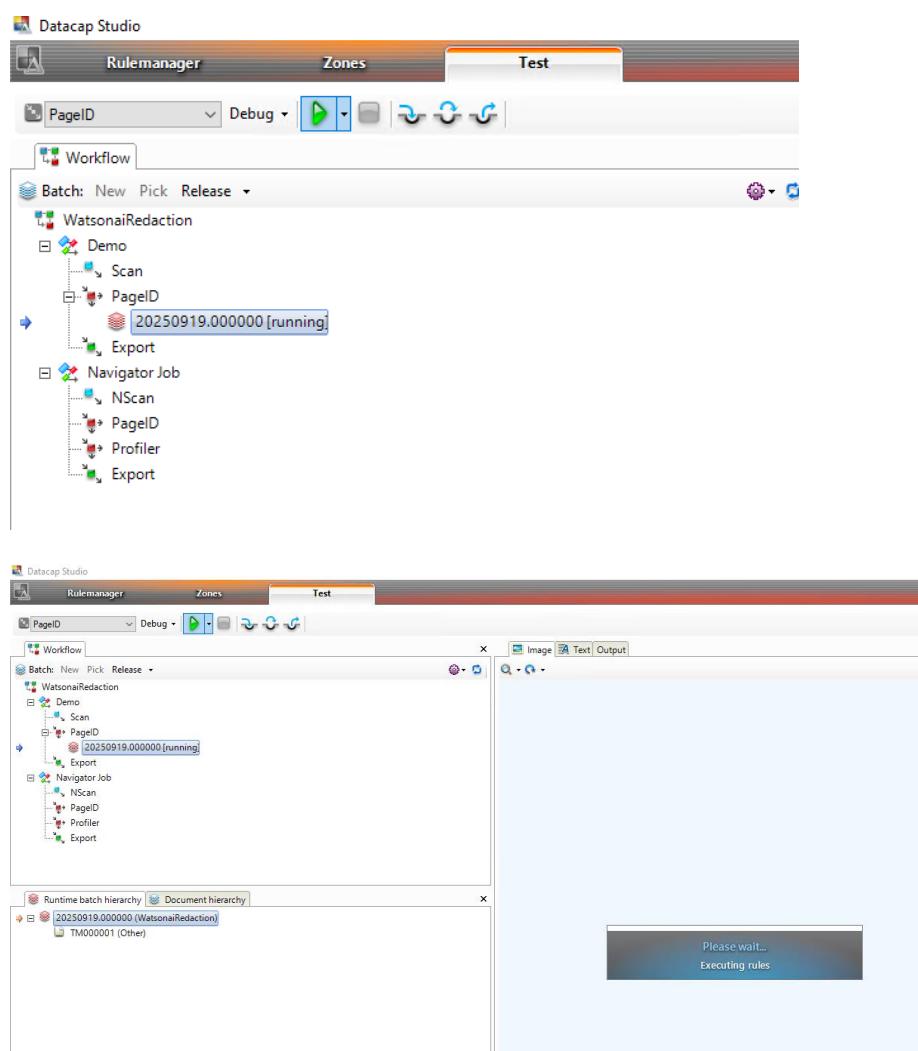
\_4. Explore the rulesets, rules, and functions assigned to the task profiles (**PageID**, **Profiler**, **Export**) and their mappings to the DCO.

\_5. Click on the **Test** tab to run the application after familiarizing yourself with how the application is created and how the workflow tasks are structured.

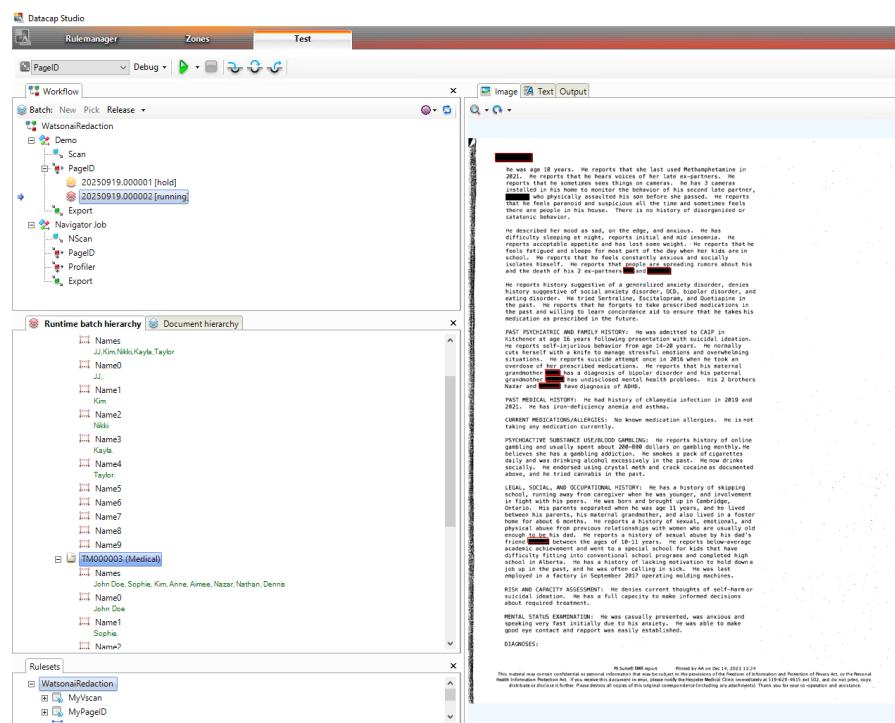
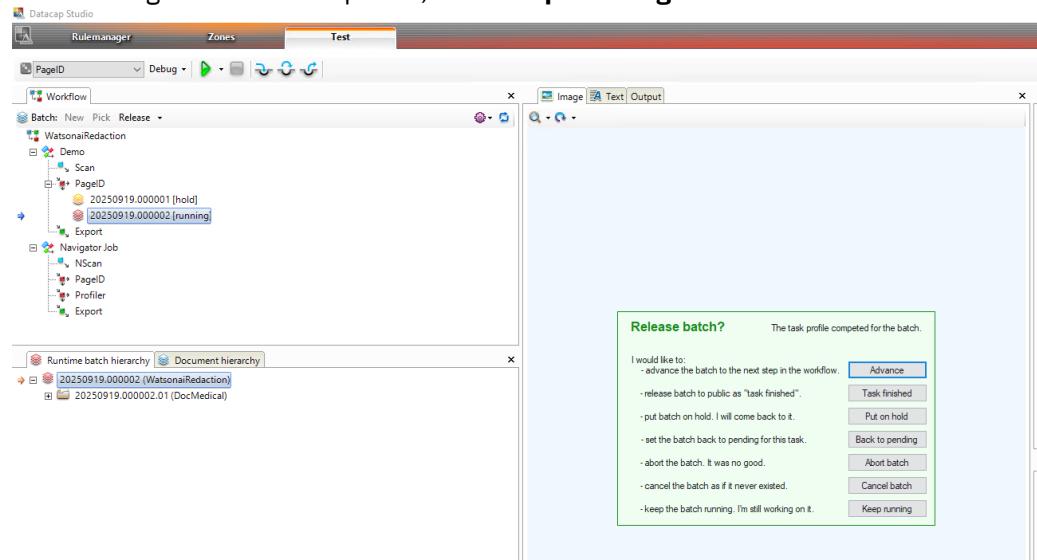
\_6. Select **Scan** from the dropdown menu, then right-click on **Scan** and choose **New** and click green play button.



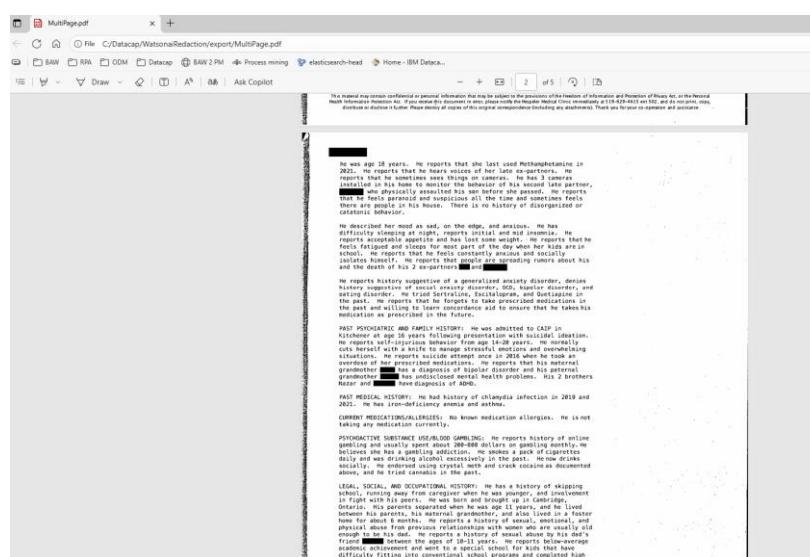
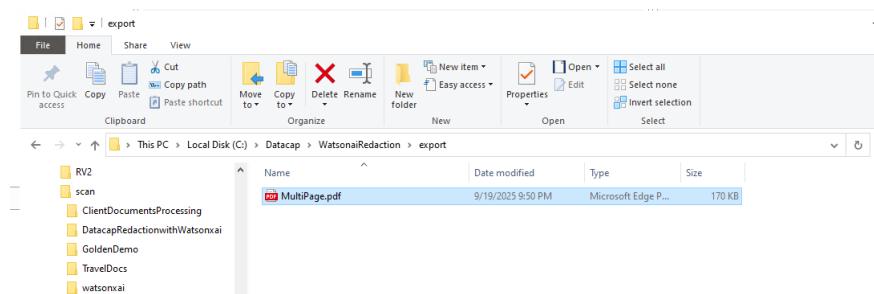
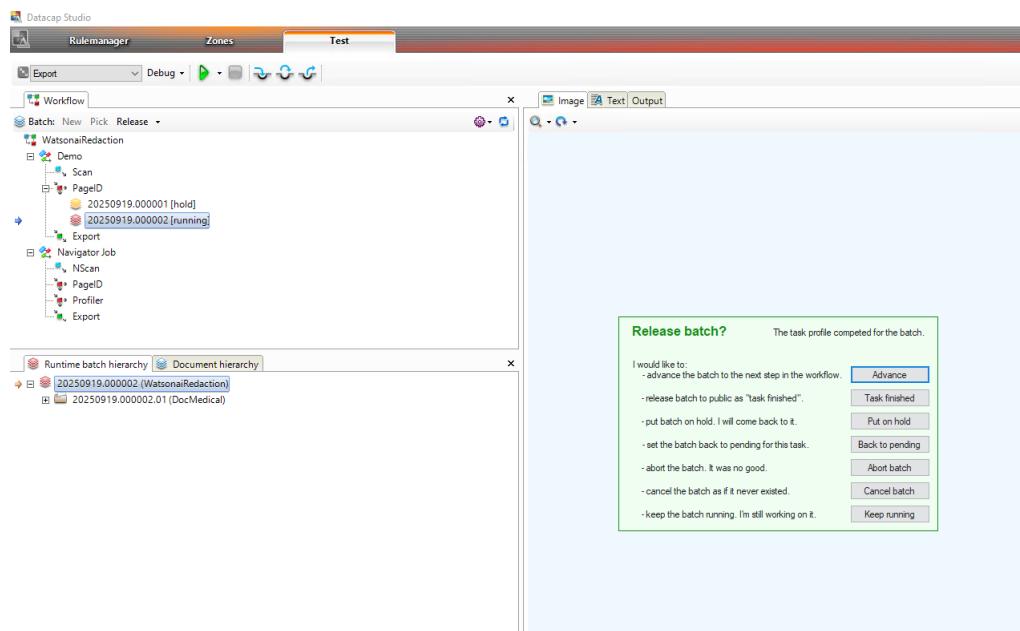
\_7. Once the scan task is completed, click **Advance** to move to the PageID task, where recognition, classification, and redaction are performed on the sample document; then, click the green **Play** button to start the process.



\_8. Once the PageID task is completed, click **Keep Running** for the moment to continue and view the document.



\_9. Once you have completed the verification, select **Export** from the dropdown menu and click the **Play** button to execute the export ruleset.



## **6. Reference Links:**

[watsonx.ai actions - IBM Documentation](#)  
[AskAFreeFormQuestion - IBM Documentation](#)  
[Datacap watsonx.ai Actions Best Practices](#)

**Congratulations you have successfully completed IBM Datacap+Watsonx.ai lab.  
We hope you enjoyed it.**