

Course Guide

Developing Applications in IBM Datacap

V9.1.7

Course code WF318G ERC 1.0



June 2020 edition

Notices

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

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*

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 jurisdictions 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 provide 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 actual people or business enterprises 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.

© 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

Trademarks	x
Course description	xi
Agenda	xiii
Unit 1. Introducing IBM Datacap	1-1
Unit objectives	1-2
Topics	1-3
1.1. Datacap overview	1-4
Datacap overview	1-5
Content is critical to every business	1-6
Benefits of a Datacap solution	1-7
What is IBM Datacap?	1-9
Structured documents	1-10
Semi-structured documents	1-11
Unstructured documents	1-12
Datacap recognizes different types of data	1-13
Highly structured data	1-14
Summary of type of documents and Datacap methods	1-15
1.2. Datacap process	1-16
Datacap process	1-17
Datacap capture process	1-18
Workflow tasks in the Datacap Navigator tool	1-19
Ingest tasks - NScan or VScan	1-20
Captures documents from many sources	1-21
Datacap supported file formats for various sources	1-22
Classification and Recognition - PagelD	1-23
Page identification by using barcodes	1-24
Sample document for page identification with barcodes	1-25
Page identification by using keywords	1-26
Sample document for page identification with keywords	1-27
Page identification by using fingerprints	1-28
Sample document for page identification with Fingerprints	1-30
Automatic Recognition from Images	1-31
Page identification with pattern recognition	1-32
Page identification with separator sheets	1-33
Extraction	1-34
Data verification	1-36
Data export	1-37
Manual and automated steps in Datacap batch	1-38
1.3. Datacap architecture	1-39
Datacap architecture	1-40
Single system configuration	1-41
Multiple-system configuration	1-42
Server client tier configuration	1-43
Components – Datacap Web Services	1-45
Components – Datacap Server	1-46
Components – Rulerunner Server	1-47
The Datacap.xml file	1-48

Unit summary	1-49
Review questions	1-50
Review answers	1-51
Unit 2. Introducing Datacap clients for processing batches	2-1
Unit objectives	2-2
Topics	2-3
2.1. Datacap Desktop	2-4
Datacap Desktop	2-5
Datacap Desktop capabilities overview	2-6
Use Datacap Desktop with your Datacap applications	2-7
Datacap Desktop – Select applications and tasks	2-9
Datacap Desktop - Monitor view	2-10
Select columns and apply filter in the Monitor view	2-11
2.2. Datacap Navigator	2-12
Datacap Navigator	2-13
What is Datacap Navigator?	2-14
Datacap View for the business users	2-15
Views in Datacap Navigator	2-17
Job Monitor	2-18
Scan Batches	2-19
Rescan options in the Scan tab	2-20
Verify batches	2-21
User settings	2-23
Unit summary	2-24
Exercise: Processing Datacap batches	2-25
Exercise objectives	2-26
Unit 3. Introducing Datacap Studio and application design.....	3-1
Unit objectives	3-2
Topics	3-3
3.1. Introduction to Datacap Studio	3-4
Introduction to Datacap Studio	3-5
Datacap Studio Overview	3-6
Connection wizard in Datacap Studio	3-7
Datacap application wizard in Datacap Studio	3-8
3.2. Datacap Studio - Rulemanager	3-9
Datacap Studio - Rulemanager	3-10
Rulemanager view	3-11
Document Hierarchy (DCO) - document and page types	3-12
Document Hierarchy (DCO) - fields	3-14
Rulesets, rules, functions, and actions	3-15
Task Profiles	3-17
Rulesets and task profiles	3-19
Actions Library	3-20
Properties tab	3-22
Rule object properties tab	3-24
Properties tab	3-26
Action object properties tab	3-28
3.3. Datacap Studio - Zones	3-29
Datacap Studio - Zones	3-30
Zones view	3-31
Fingerprints pane	3-33
Document hierarchy pane	3-34
Image View	3-35
Zones	3-36

3.4. Datacap application design	3-37
Datacap application design	3-38
Steps to design and configure an application	3-39
Do a complete end-to-end evaluation	3-40
Scenario for a sample application	3-41
Deciding which method to use for document classification	3-42
Planning the workflow details for the application	3-43
Planning the task details for the workflow	3-44
Steps to build an application	3-45
Set up Document hierarchy in Datacap Studio	3-46
Order of Operations in Datacap	3-47
Rulesets and Task profiles for the application	3-48
Example of Page ID and document classification	3-49
Unit summary	3-50
Review questions	3-51
Review answers	3-52

Unit 4. Building a Datacap application with Forms Template 4-1

Unit objectives	4-2
Topics	4-3
4.1. Create a Datacap application with Forms Template.	4-4
Create a Datacap application with Forms Template	4-5
Scenario for the application	4-6
Steps to build an Datacap application	4-7
Datacap application templates	4-8
Template Folder Structure	4-9
Create an application	4-10
Datacap.xml	4-12
4.2. Configure the application in IBM Content Navigator.	4-13
Configure the application in IBM Content Navigator	4-14
IBM Content Navigator administration - create a repository	4-15
Add a repository to the Datacap user desktop	4-16
Configure jobs for a workflow in Datacap Admin Console	4-17
Datacap Admin Console - add a task for a job	4-18
Datacap Admin Console - edit a task for a job	4-19
4.3. Edit the application settings in Datacap Application Manager	4-20
Edit the application settings in Datacap Application Manager	4-21
Application Manager settings	4-22
Application Manager settings	4-23
4.4. Define document hierarchy.	4-24
Define document hierarchy	4-25
Configuring Datacap application in Datacap Studio	4-26
Define Document Hierarchy (DCO) objects	4-27
Associate rulesets with Document hierarchy (DCO) objects	4-29
4.5. Configure fingerprints and zones	4-30
Configure fingerprints and zones	4-31
Fingerprints	4-32
Page identification – Fingerprints	4-33
Configure Fingerprints in the Zones tab	4-34
Configure Zones	4-35
Set OCR filters for the fields	4-36
Unit summary	4-37
Review questions	4-38
Review answers	4-39
Exercise: Building a Datacap application by using Forms Template	4-40
Exercise objectives	4-41

Unit 5. Configuring rulesets for a Datacap application.....	5-1
Unit objectives	5-2
Topics	5-3
5.1. Add rulesets	5-4
Add rulesets	5-5
Rulesets, rules, functions, and actions	5-6
List of rulesets and task profiles that are used for the application	5-8
Toolbar in Datacap Studio Rulesets pane	5-10
Add an action to a function	5-11
Set parameters to an action	5-12
VScan ruleset	5-13
Page ID ruleset	5-14
CreateDocs&Fields ruleset	5-15
Recognize and Validate rulesets	5-16
FingerprintAdd ruleset	5-17
Export the data	5-18
5.2. Configure Task profiles.....	5-19
Configure Task profiles	5-20
Add a ruleset to a task profile	5-21
5.3. Test the application.....	5-22
Test the application	5-23
Testing the application at design time	5-24
Testing the application - Advance the batch	5-25
Testing the application - Page ID and Profiler tasks	5-26
Testing the application - Export task	5-27
Unit summary	5-28
Review questions	5-29
Review answers	5-30
Review answers	5-31
Exercise: Configuring rulesets for a Datacap application	5-32
Exercise objectives	5-33
Unit 6. Troubleshooting a Datacap application.....	6-1
Unit objectives	6-2
Topics	6-3
6.1. Examine the logs files for debugging	6-4
Examine the logs files for debugging	6-5
Logs files that are used in debugging	6-6
Sample log file - task completed successfully	6-7
Sample log file - task was not completed successfully	6-8
6.2. Troubleshoot the applications in the Datacap Studio Test tab	6-9
Troubleshoot the applications in the Datacap Studio Test tab	6-10
Testing the application at design time	6-11
Datacap Studio Test tab debugging features	6-12
Breakpoint pane in the Datacap Studio Test tab	6-13
Setting generic breakpoints	6-14
Setting specific breakpoints on batch hierarchy	6-15
Setting specific breakpoints on rulesets, rules, or actions	6-16
Disabling and clearing breakpoints	6-17
UI controls for Stepping through the code	6-18
Single stepping through a batch	6-19
Examining log files from the Test tab	6-20
Unit summary	6-21
Review questions	6-22
Review answers	6-23
Exercise: Troubleshooting a Datacap application	6-24

Exercise objectives	6-25
Unit 7. Building an application to process multiple page types	7-1
Unit objectives	7-2
Topics	7-3
7.1. Copy an existing application	7-4
Copy an existing application	7-5
Scenario for the application	7-6
Steps to configure the Datacap application	7-7
Copy an existing application to create a new application	7-8
Datacap.xml	7-9
7.2. Configure the application in IBM Content Navigator	7-10
Configure the application in IBM Content Navigator	7-11
Create a repository in IBM Content Navigator	7-12
Add a repository to the Datacap admin desktop	7-13
Select your application in Datacap Administration Console	7-14
Rename a job in Datacap Admin Console	7-15
7.3. Define document hierarchy and configure rulesets	7-16
Define document hierarchy and configure rulesets	7-17
Define Document hierarchy objects	7-18
List of rulesets and task profiles for the application	7-20
VScan ruleset	7-21
Page ID ruleset	7-22
CreateDocs&Fields ruleset	7-24
Recognize and Validate rulesets	7-25
Fingerprint ruleset	7-26
Export the data	7-27
7.4. Configure fingerprints and zones	7-29
Configure fingerprints and zones	7-30
Fingerprints	7-31
Configure fingerprints in the Zones tab	7-32
Configure zones	7-33
7.5. Test the application	7-34
Test the application	7-35
Testing the application at design time	7-36
Testing the application - Page ID and Profiler tasks	7-37
Testing the application - Export task	7-38
Unit summary	7-39
Review questions	7-40
Review answers	7-41
Exercise: Building an application to process multiple page types	7-42
Exercise objectives	7-43
Unit 8. Configuring an application for recognition with OMR	8-1
Unit objectives	8-2
Topics	8-3
8.1. Configure the application for recognition with OMR	8-4
Configure the application for recognition with OMR	8-5
Optical Mark Recognition (OMR)	8-6
Define a dictionary	8-7
Configure variables for an OMR field	8-8
Define document hierarchy for an OMR field	8-9
Configure the CreateDocs&Fields ruleset	8-10
Configure the Recognize ruleset and zones for the field	8-11
8.2. Identify the batch results output	8-12
Identify the batch results output	8-13

Testing the application - Page ID task	8-14
Testing the application - Profiler task	8-15
Testing the application - Export task without a dictionary	8-16
Testing the application - Export task with a dictionary	8-17
Unit summary	8-18
Review questions	8-19
Review answers	8-20
Exercise: Configuring an application for recognition with OMR	8-21
Exercise objectives	8-22
Unit 9. Exporting data to a FileNet Content Manager repository.....	9-1
Unit objectives	9-2
Topics	9-3
9.1. Export data to a FileNet Content Manager repository	9-4
Export data to a FileNet Content Manager repository	9-5
Export of scanned documents and data	9-6
Preparation for export to a FileNet repository: Document	9-8
Preparation for export to a FileNet repository: Folder	9-9
Export to FileNet P8 ruleset: Set parameters	9-10
Export: Set document parameters and map property values	9-11
Export to FileNet P8 ruleset: Document upload	9-12
Preparing the data for the repository upload	9-13
9.2. Verify the exported documents	9-14
Verify the exported documents	9-15
Verifying the export on the repository	9-16
Unit summary	9-17
Review questions	9-18
Review answers	9-19
Exercise: Exporting data to a FileNet Content Manager repository	9-20
Exercise objectives	9-21
Unit 10. Creating page layouts	10-1
Unit objectives	10-2
Topics	10-3
10.1. Introduction to Datacap Insight Edition	10-4
Introduction to Datacap Insight Edition	10-5
Unstructured documents	10-6
IBM Datacap Insight Edition	10-7
10.2. Creating page layouts	10-8
Creating page layouts	10-9
Page layout XML file	10-10
The PDFFREDocumentToImage action	10-11
The Recognize action	10-12
Recognizing the PDF and the Image files	10-13
Analyze layout	10-14
Populate Layout field in DCO	10-16
Test the application	10-17
Unit summary	10-18
Review questions	10-19
Review answers	10-20
Exercise: Creating page layouts	10-21
Exercise objectives	10-22
Unit 11. Creating a virtual page block and extracting data from tables	11-1
Unit objectives	11-2
Topics	11-3

11.1.	Create a virtual page block	11-4
	Create a virtual page block	11-5
	What is a virtual page block (VPB)?	11-6
	Sample ruleset to create a virtual page block (VPB)	11-7
	Verifying a virtual page block (VPB)	11-8
11.2.	Configure data extraction from tables	11-9
	Configure data extraction from tables	11-10
	Sample ruleset for Analyze layout for VPB	11-11
	The FindTableValueRegEx action	11-13
	Test the table extraction	11-15
	Export the table data	11-16
	Unit summary	11-17
	Review questions	11-18
	Review answers	11-19
	Exercise: Creating a virtual page block and extracting values from tables	11-20
	Exercise objectives	11-21
Unit 12.	Extracting data from label-value pairs	12-1
	Unit objectives	12-2
	Topics	12-3
12.1.	Configure the Datacap application for data extraction from label-value pairs	12-4
	Configure the Datacap application for data extraction from label-value pairs	12-5
	Sample scan image for label-value pairs	12-6
	Find the label-value pairs from the layout XML file	12-7
	Test the Find label value pairs configuration	12-8
	Extract the data from a specific label-value pair	12-9
	Test the data extraction for specific label-value pairs	12-10
	Export the data from label-value pairs	12-11
	Unit summary	12-12
	Review questions	12-13
	Review answers	12-14
	Exercise: Extracting data from label-value pairs	12-15
	Exercise objectives	12-16
Unit 13.	Course summary, badge, and other learning resources	13-1
	Unit objectives	13-2
	Course objectives (1 of 2)	13-3
	Course objectives (2 of 2)	13-4
	IBM badge	13-5
	IBM Professional Certifications	13-6
	Other learning resources (1 of 4)	13-7
	Other learning resources (2 of 4)	13-8
	Other learning resources (3 of 4)	13-9
	Other learning resources (4 of 4)	13-10
	Unit summary	13-11
	Course completion	13-12

Trademarks

The reader should recognize that the following terms, which appear in the content of this training document, are official trademarks of IBM or other companies:

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide.

The following are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide:

FileNet®

Insight®

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

Social® is a trademark or registered trademark of TWC Product and Technology, LLC, an IBM Company.

Other product and service names might be trademarks of IBM or other companies.

Course description

Developing Applications in IBM Datacap V9.1.7

Duration: 4 days

Purpose

This course provides technical professionals with the skills that are needed to build Datacap applications.

The course begins with an introduction to IBM Datacap. You learn about capture concepts, Datacap process, page identification methods, and architecture. You process batches for Datacap applications in the Datacap clients.

You learn about the design and components of a Datacap application. You build a Datacap application by using Forms Template in Datacap Studio and configure it. You learn how to troubleshoot a Datacap application. You configure a Datacap application to process documents of multiple page types in a single batch. You implement OCR and OMR to extract data from data fields and from multiple choice check boxes. You export data to a text file and also to an IBM FileNet Content Manager repository.

You build page layouts, create virtual page blocks, and extract data from tables and label-value pairs.

Through instructor-led presentations and hands-on lab exercises, you learn about the core features of IBM Datacap.

Audience

This course is intended for application developers of IBM Datacap.

Prerequisites

- Familiarity with data capture concepts

Objectives

- Describe capture concepts and Datacap
- Identify the tasks in Datacap process
- Describe the Datacap recognition methods that are used for scanned images
- Understand Datacap architecture
- Process batches in the Datacap clients
- Identify the application development features in Datacap Studio
- Describe the Datacap application design concepts

- Build a Datacap application with Forms Template
- Configure rulesets for a Datacap application
- Examine the logs files for debugging
- Troubleshoot the applications in the Datacap Studio Test tab
- Build an application for processing multiple page types in a batch
- Configure an application for recognition with OMR
- Export data and scanned documents to a FileNet Content Manager repository
- Configure a Datacap application to create page layouts
- Create a virtual page block
- Extract data from tables and label-value pairs

Contents

Refer to the contents section (TOC) for course content.

Agenda



Note

The following unit and exercise durations are estimates, and might not reflect every class experience.

Day 1

- (00:15) Course introduction
- (01:30) Unit 1. Introducing IBM Datacap
- (00:30) Unit 2. Introducing Datacap clients for processing batches
- (01:45) Exercise 1. Processing Datacap batches
- (01:30) Unit 3. Introducing Datacap Studio and application design
- (00:45) Unit 4. Building a Datacap application with Forms Template

Day 2

- (02:00) Exercise 2. Building a Datacap application with Forms Template
- (00:45) Unit 5. Configuring rulesets for a Datacap application
- (02:00) Exercise 3. Configuring rulesets for a Datacap application
- (00:30) Unit 6. Troubleshooting a Datacap application
- (01:00) Exercise 4. Troubleshooting a Datacap application

Day 3

- (00:40) Unit 7. Building an application to process multiple page types
- (02:15) Exercise 5. Building an application to process multiple page types
- (00:30) Unit 8. Configuring an application for recognition with OMR
- (01:30) Exercise 6. Configuring an application for recognition with OMR
- (00:30) Unit 9. Exporting data to a FileNet Content Manager repository
- (01:15) Exercise 7. Exporting data to a FileNet Content Manager repository

Day 4

- (00:20) Unit 10. Creating page layouts
- (01:00) Exercise 8. Creating page layouts
- (00:30) Unit 11. Creating a virtual page block and extracting values from tables
- (02:00) Exercise 9. Creating a virtual page block and extracting values from tables
- (00:30) Unit 12. Extracting data from label-value pairs
- (01:30) Exercise 10. Extracting data from label-value pairs

Unit 1. Introducing IBM Datacap

Estimated time

01:30

Overview

This unit introduces core concepts of capture and Datacap.

How you will check your progress

- Review

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Describe capture concepts and Datacap
- Identify the tasks in Datacap process
- Understand Datacap Architecture
- Identify Datacap service components

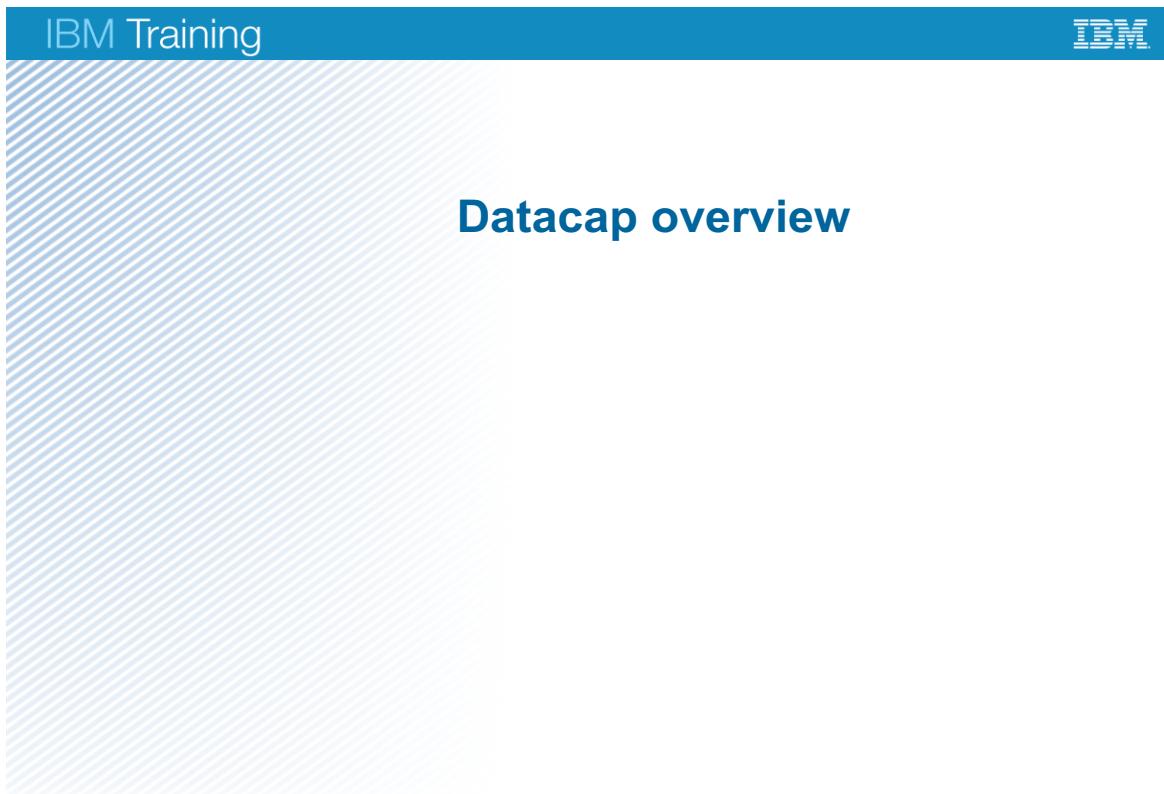
Figure 1-1. Unit objectives

Topics

- Datacap overview
- Datacap process
- Datacap architecture

Figure 1-2. Topics

1.1. Datacap overview



Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-3. Datacap overview

Content is critical to every business

 Banking	 Government	 Insurance
<ul style="list-style-type: none"> • Account opening • Lending applications • Investment reports • Regulatory, Compliance 	<ul style="list-style-type: none"> • Benefits management • Social services • Permits and licenses • Program enrollment • Records Management 	<ul style="list-style-type: none"> • Claims Processing • Underwriting • Enrollment • Appraisals • Investigations
 Healthcare	 Energy & Utilities	Cross Industry
<ul style="list-style-type: none"> • Medical claims • Patient Onboarding and discharge • Prescription Processing • Health record 	<ul style="list-style-type: none"> • Account Opening • Permits & Planning • Correspondence • Applications • Regulatory, Compliance 	<ul style="list-style-type: none"> • Accounts Payable • Invoice processing • Human resources • Licensing & Certification

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-4. Content is critical to every business

Every Industry has important business documents that need to be captured.

- Some processes are common across sectors such as on-boarding an employee, or a new customer. You need contact information and personally sensitive information such as Social Security Number (SSN), employee number, account number. You might also need contracts and signatures.
- Other sectors have special requirements. For example, financial services and government have specific records-keeping requirements such as how they manage documents and for how long.
- Across all sectors, organizations are struggling to deal with documents to unlock the valuable information inside the PDF files, emails, faxes, and on paper, and to be able to activate information.

Benefits of a Datacap solution

	Cost Savings	<ul style="list-style-type: none"> ▪ Automation of processes by using capture can deliver 50% - 70% savings
	Productivity	<ul style="list-style-type: none"> ▪ Automation tools can be up to 5X faster than human resource ▪ Can function 24x7 ▪ Enables people to focus on more important tasks
	Accuracy	<ul style="list-style-type: none"> ▪ Accuracy in processing can reach 90+%
	Compliance	<ul style="list-style-type: none"> ▪ Geographic requirements can be accommodated. ▪ Rules driven transaction processing is consistently applied for fiscal requirements
	Scalability and flexibility	<ul style="list-style-type: none"> ▪ Capture tools can be adapted to multiple geos and processes. ▪ Can scale to accommodate seasonality

Figure 1-5. Benefits of a Datacap solution

IBM Datacap is designed for enterprise-wide deployments in paper-intensive market segments, such as government, insurance, healthcare, financial services, and transportation, to name a few.

Cost savings

IBM Datacap helps reduce the cost through automation of the ingestion of documents, extraction of data, and export of data and documents to make them available for business processes.

Distributed enable clients to reduce or eliminate document shipping costs and distribute labor to areas with more affordable labor rates.

Productivity

By significantly reducing manual data entry and paper-based storage and retrieval of documents, knowledge workers can more quickly process and access documents. Quicker access to documents improves customer service case management, business transactions, and compliance.

Accuracy

By minimizing the human data entry errors, IBM Datacap can improve accuracy of data. Time is saved tracking down misplaced document images.

Streamlined process: IBM Datacap eliminates a cumbersome paper process by enabling clients to automate previously labor-intensive aspects.

What is IBM Datacap?

- Datacap:
- Is an advanced Document Imaging and capture software
- Automates capturing documents
- Automates the extraction of appropriate data
- Can handle different types of documents regardless of what business processes are involved
 - Structured documents
 - Semi-structured documents
 - Unstructured documents
- Exports data to different repositories
- Pass the data to other applications
- Offers customizable solutions

Figure 1-6. What is IBM Datacap?

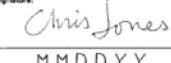
IBM Datacap is an advanced Document Imaging and capture software. It automates the capturing documents and extraction of appropriate data.

Automation accelerates the process, and business can respond quickly to their customers and Business Partners. The extracted data can be stored in multiple ways for access. You can also pass the data to other applications.

Structured documents

- Structured documents
 - All documents have a consistent format
 - Form taxonomy doesn't change
 - Every document has the same data fields in the same place
- Examples
 - Tax forms
 - Beneficiary forms
 - Claims
 - Reimbursement forms
- Datacap method
 - Fingerprinting

Auto Insurance Claim Form
Insurance Company A

Policy Holder Address :	Chris Jones, 13 Johan Street NY 987654
Driver Name :	Chris Jones
Policy Number :	48998813
Incident Number :	CL-4328941
Incident Date:	03/03/11
Incident Time:	14:15
Vehicle License :	2296RG
Vehicle Colour :	Black
Vehicle Manufacturer :	Toshiba
Vehicle Model :	Quaser
Year of Manufacture :	2009
Chassis Number (VIN) :	2C5BB4CRX82LP3489
Incident Description: Incident occurred on Highway 69, NY. Damage occurred to rear of the vehicle. Damaged right tail light and bumper.	
Was anyone injured in this incident which required medical attention? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No	
By signing the adjacent box and entering the current date, you agree the information above is accurate to the best of your knowledge. Signature:  M M D D Y Y Dated: 03/03/11	

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-7. Structured documents

The screen capture shows an example of a structured form.

When an organization owns the documents, they can maintain them as structured documents with predictable data fields.

Traditional capture systems can capture known and predictable document types such as forms.

They create an image, extract values from pre-defined areas on the form, and store the data.

- Uses OCR capture and pre-defined rules
- Works well with known forms and structured documents
- Can be accurate and fast when documents are consistent and predictable

IBM Training



Semi-structured documents

- Semi-structured documents
 - All documents do not have a consistent format
 - Form taxonomy can change
- Every document does not have the same data fields in the same place
- Example:
 - Invoice
- Datacap methods
 - Locate action library
 - Text analytics

B	Brilliant Repair Shop 90 Fixed Street Carson City, NV 89701 Phone: 775-555-6789	INVOICE
S		INVOICE #03-8508 DATE: 02/15/20
To: Busy Car Repair 100 Auto Road Salt Lake City, UT 84101 Phone: 801-555-1234		Ship To: Busy Car Repair 100 Auto Road Salt Lake City, UT 84101 Phone: 801-555-1234
Comments or special instructions:		
SALESPERSON 198	PO. NUMBER 0811012	REGISTRATION SHIPPED VIA FO.B. POINT TERMS 2/10 Net 30
QUANTITY	DESCRIPTION	UNIT PRICE
1	Left quarter panel	175.00
2	Tail lights	50.00
80	Gaskets	.95
2	Front headlights	50.00
80	Bolts	.95
		TOTAL
		175.00
SUBTOTAL		572.00
SALES TAX		27.68
SHIPPING & HANDLING		
TOTAL DUE		554.68
Make all checks payable to Brilliant Repair Shop. If you have any questions concerning this invoice, contact (775) 555-6789.		
Thank you for your business!		

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-8. Semi-structured documents

The screen capture shows an example of a semi-structured form.

The types of documents that come to a company from other vendors can be different. For example, different types of invoices.

Also, in the same invoice format, depending on how many items are listed, the “total” field can be at a different position on each Invoice.

You can use Datacap Locate action library and Text analytics to extract data from semi-structured documents.



Unstructured documents

- Documents do not have a predictable format. For example, Letters, Contracts
 - Datacap offers Insight edition tools to handle unstructured documents
 - Datacap solution needs to be customized

BILLS OF LADING	DRAFTS	PACKING LISTS
CERTIFICATES		
BILL OF EXCHANGE		
LEASES		
CONTRACTS		

Figure 1-9. Unstructured documents

The screen captures show examples of unstructured forms.

Most of the important information that comes to an organization is not on a form but is in a letter, an email message or in a PDF file. The traditional document capture systems cannot handle that sort of unstructured data.

IBM Datacap uses cognitive technologies such as Natural Language Processing to help change the approach and understanding on the unstructured documents. It can ingest this unstructured data, understand it, reason about it, and learn from it. It can look at the structure and content to determine a document type, and then use text analytics to identify entities, patterns, and keywords to extract important information.

The cognitive capture automates the processing of document transactions by using advanced document imaging, Content Analytics, and machine learning techniques to process all document types.



Datacap recognizes different types of data

- Key-value pair
- Hand-writing
- Checkboxes
- Highly structured data

Rental Agreement Number: 703142974

Return Date/Time: 23APR10/1551

Handwriting
Canyon Country

Was anyone injured in this incident which required medical attention?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No
---	------------------------------	--

Figure 1-10. Datacap recognizes different types of data

Datacap can recognize different types of data.

In a key-value pair, for example, Rental Agreement Number: 703142974, the key is Rental Agreement Number and the value is 703142974. In a key-value pair, a colon after the key is optional.

In Datacap, you can use the following methods to extract different data types:

- Fingerprinting and zoning for the fields in structured forms
- Locate actions for the fields in semi-structured forms
- OMR for check boxes
- Text analytics for unstructured forms

Next slide discusses the highly structured data.

Highly structured data

- Data with specialized formats
- Can be identified by their structure, even if they are presented without any context
- Examples:
 - Date: 15 March 2020
 - Phone number: (555)555-5555
 - Email: developer@datacap.com
- Without zoning or fingerprinting, you can easily locate the data
- Datacap offers Insight edition tools to handle highly structured data

Figure 1-11. Highly structured data

For the highly structured data, you probably cannot use zoning because the location of the data changes.

You can identify the structures for highly structured data. You might still want to use label-value pair extraction. For example, to differentiate between a check-in date versus check-out date. Similarly, a dollar amount might be recognizable. You still need to differentiate if there is a tax amount and a total amount fields.

Summary of type of documents and Datacap methods

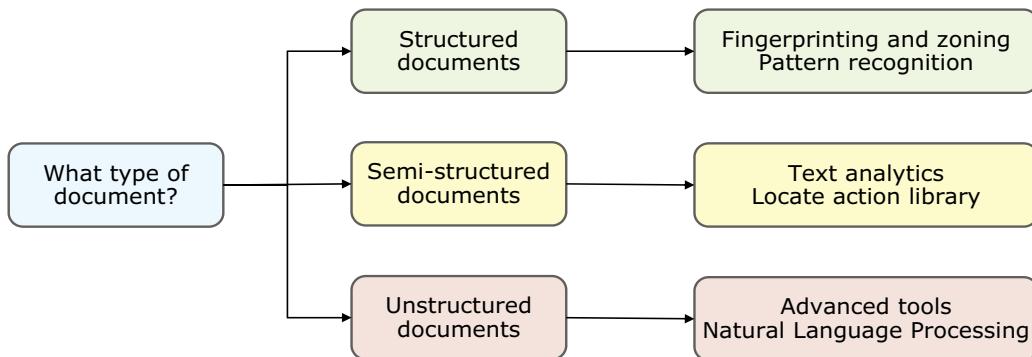
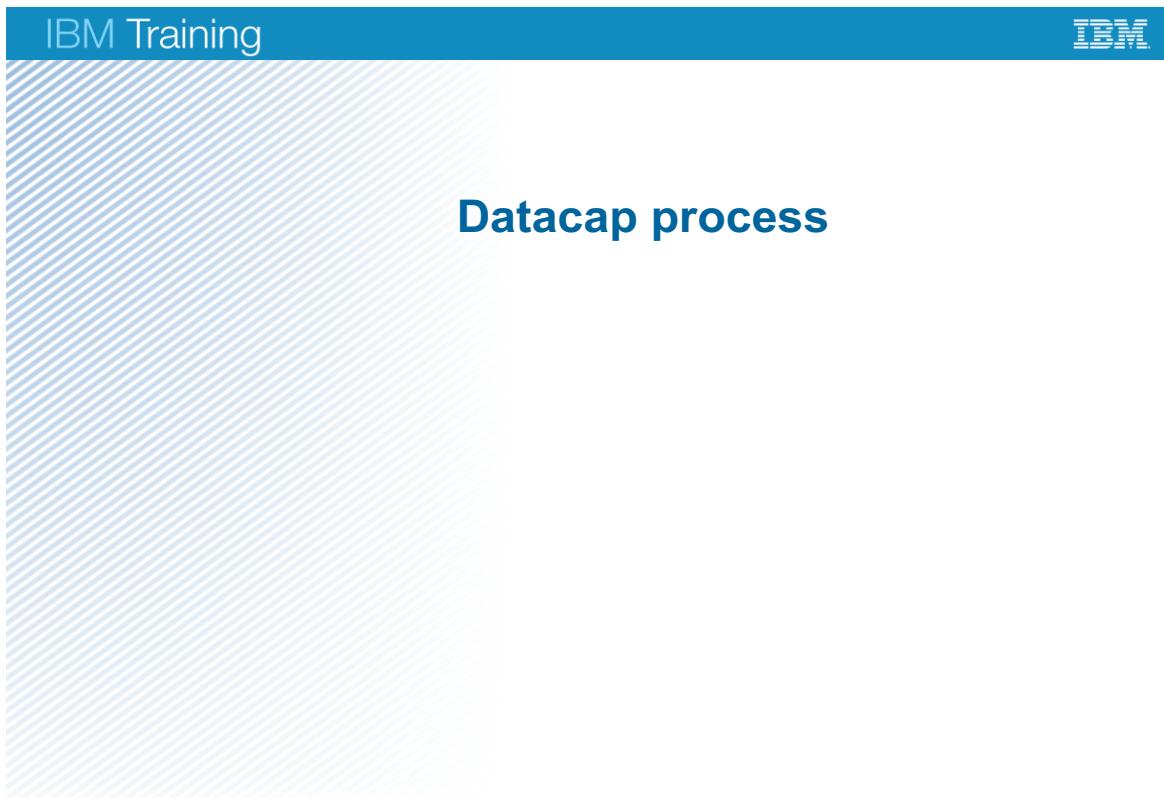


Figure 1-12. Summary of type of documents and Datacap methods

The screen capture shows the type of documents types and the Datacap solutions that can be used.

- Structured documents
Fingerprinting, zoning and pattern recognition
- Semi-structured documents
Text analytics and Locate action library
- Unstructured documents
Advanced tools, Natural Language Processing

1.2. Datacap process



Introducing IBM Datacap

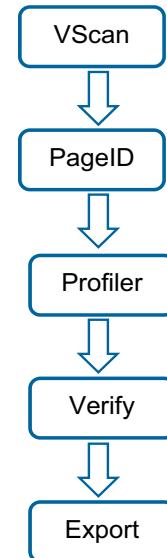
© Copyright IBM Corporation 2020

Figure 1-13. Datacap process

Datacap capture process

Datacap Workflow tasks:

1. Ingest documents
NScan or VScan
2. Classify and recognition
PageID
3. Extract data
Profiler
4. Verify or double blind verify
Verify
5. Pre-Export & Export
Export



[Introducing IBM Datacap](#)

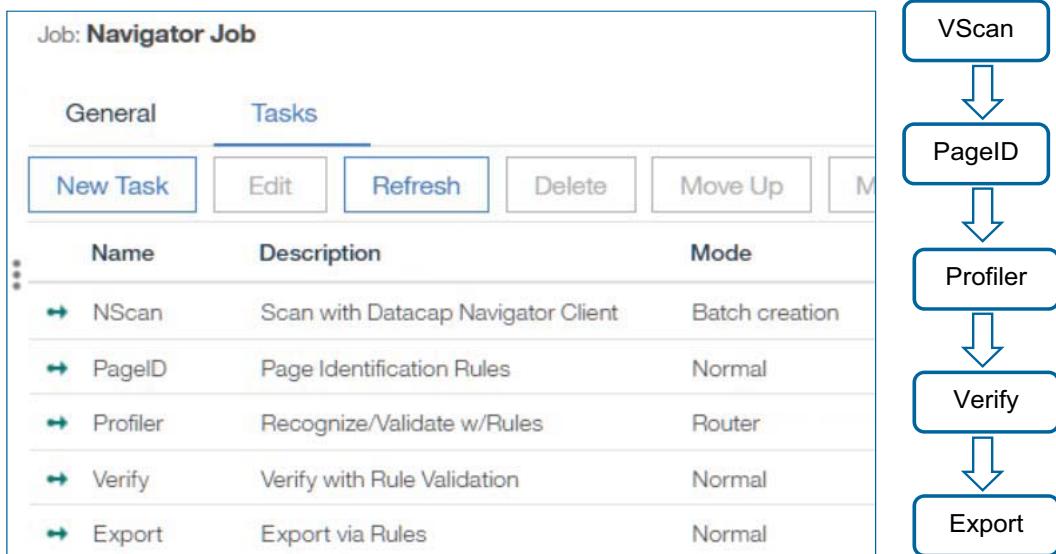
© Copyright IBM Corporation 2020

Figure 1-14. Datacap capture process

The screen capture shows the Datacap workflow tasks.

You can organize the flow of tasks in the capture process from scan to export and exceptions in a workflow.

Workflow tasks in the Datacap Navigator tool



[Introducing IBM Datacap](#)

© Copyright IBM Corporation 2020

Figure 1-15. Workflow tasks in the Datacap Navigator tool

The screen capture shows a sample list of workflow tasks for a Datacap application in the Datacap Navigator administration tool.

You can configure the Datacap tasks in the Datacap Navigator administration tool.

Ingest tasks - NScan or VScan

- Naming convention for ingest tasks:
 - Standard - VScan
 - Datacap Navigator - NScan
 - For emails - Email
- Sources for capturing
 - Scanners
 - Fax machines
 - Mobile
 - Network folders
 - Multi-functional devices
 - Emails
- Setup workflow tasks based on the source

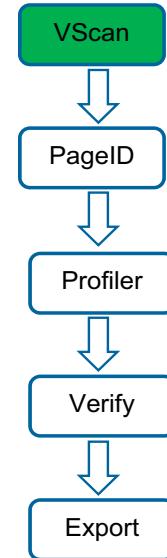


Figure 1-16. Ingest tasks - NScan or VScan

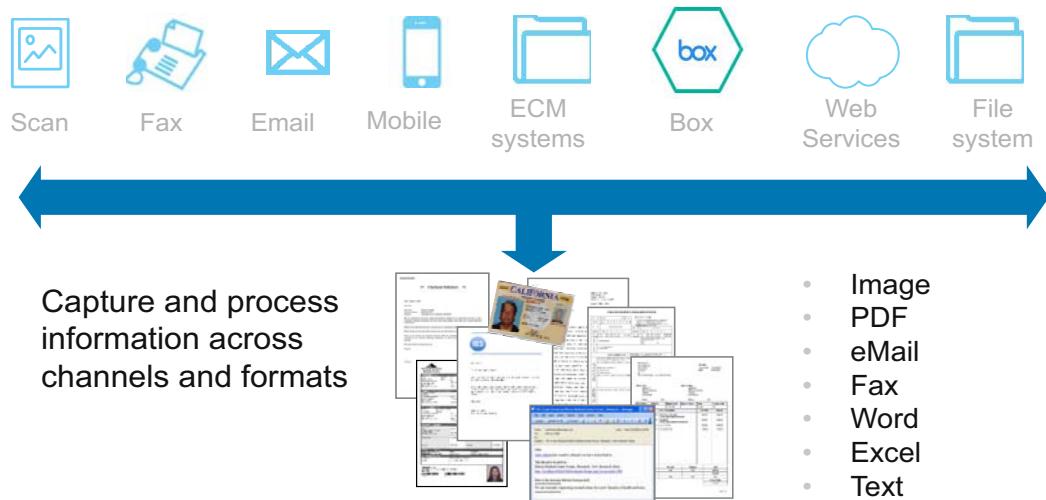
The diagram on the slide shows the flow of tasks for a Datacap application.

You can scan the documents either in the Datacap Navigator (web client) or in the Datacap Desktop (thick) clients.

IBM Training



Captures documents from many sources



Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-17. Captures documents from many sources

Datacap can capture and process information across the following channels:

- Scanner and Fax systems
- Email
- Mobile
- ECM systems
- Box
- Web services
- File systems

Datacap can capture and process information that of following formats:

- Image
- PDF and Text
- Email
- Fax
- Word and Excel

Datacap supported file formats for various sources

Input Channels	File Format	Notes
<ul style="list-style-type: none"> • Scanners • Multi-Functional Devices 	<ul style="list-style-type: none"> • TIFF, JPEG, PDF 	Color and grayscale are typically converted to single page bitonal TIFF.
<ul style="list-style-type: none"> • Mobile devices • Email attachments • Windows file system 	<ul style="list-style-type: none"> • TIFF, JPEG, PDF, PNG • HTML, RTF, TXT, DOC, DOCX, XLS, XLSX, ZIP, EML 	<ul style="list-style-type: none"> • JPEG2000 compression is not supported. • Password-protected PDF files and fillable PDF forms are not supported.
<ul style="list-style-type: none"> • FAX 	<ul style="list-style-type: none"> • G3, or G4 TIFF • Single-page or multi-page bitonal 	<ul style="list-style-type: none"> • For recognition and all tasks. • Fax documents are black and white.

Figure 1-18. Datacap supported file formats for various sources

The slide shows a list of channels that are sources for file input and their supported file formats.

Classification and Recognition - PageID

- Convert
 - To TIFF
 - If the file format is different
- Image Enhance
- Classify
 1. Barcodes
 2. Fingerprints or pattern recognition
 - OCR decision
 3. Text analytics (keywords)
 4. Identification by human in Datacap Navigator
- Create Docs
 - Create docs
 - Create fields

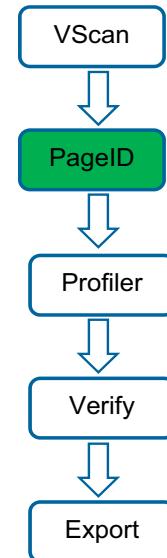


Figure 1-19. Classification and Recognition - PageID

Convert

If the file format is different from TIFF, such as PDF, Word, or PNG, the file needs to be converted into TIFF before the page identification.

Image Enhance

Use methods such as Deskew, line removal, despeckle to clean the image

By using the Datacap rules engine, data capture can be tailored to fit the most demanding business requirements.

- The settings can be changed quickly when business needs change
- Most of the documents can be recognized and classified by using automatic methods
- For a small percentage that cannot be classified by any methods, it is also possible to classify by human in Datacap Navigator

Page identification by using barcodes

- Identifies the current page based on the barcode values that are found in the image
- One- and two-dimensional barcodes are used for page recognition
- One-dimensional barcodes
 - Can be with or without numbers
 - Example: Code 39 can store up to 43 alphanumeric characters
- Two-dimensional (2D) barcodes
 - Coded with a matrix that represents information along the vertical and horizontal axes of the barcode
 - Can store up to several KB of data



Figure 1-20. Page identification by using barcodes

The diagrams on the slide show different types of barcodes.

Sample document for page identification with barcodes

- The page can be identified by using a barcode.
- The document has two barcodes:
 1. Upper barcode represents the data and it can change
 - Cannot be used for Page ID
 2. Lower barcode represents the page
 - Example: Claim - 29A
 - Data does not change
 - Can be used for identifying the page

Auto Insurance Claim Form
Insurance Company A

Policy Holder Address :	Chris Jones, 13 Johan Street NY 987654
Driver Name :	Chris Jones
Policy Number :	48998813
Incident Number :	CL-4328941
Incident Date:	03/03/11
Incident Time:	14:15
Vehicle License :	2296RG
Vehicle Colour :	Black
Vehicle Manufacturer :	Toa
Vehicle Model :	Quaser
Year of Manufacture :	2009
Chassis Number (VIN) :	2C5BB4CRX82LP3489
Incident Description: Incident occurred on Highway 69, NY. Damage occurred to rear of the vehicle. Damaged right tail light and bumper.	
Was anyone injured in this incident which required medical attention? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No	
By signing the adjacent box and entering the current date, you agree the information above is accurate to the best of your knowledge. Signature: M M D D Y Y Dated: 03/03/2011	

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-21. Sample document for page identification with barcodes

The screen capture shows an example of a structured form with a barcode that can be used for page identification.

Page identification by using keywords

- Keyword identifies the current page based on the keywords that are found in the recognition text
 - This identification technique requires recognition text and full page recognition enabled
- Word matching is case-sensitive
- Keyword text file
 - Contains list of words or phrases that are separated by new lines
 - Must have a ".key" extension for the system to recognize it as a keyword file
 - Is used for matching
- Examples of keywords:
 - Donation, invoice, receipt, ticket

Figure 1-22. Page identification by using keywords

How do keywords work in page identification?

- The search first looks for the first word or phrase in the keyword file.
- Common substitutions are applied to search criteria to improve results.
- Starting from the location of the last find, if the word is found, the search stops.
- If no match is found, the next line from the keyword file is read and again the search starts from the result of a previous find.
- This process continues until a match is found or all of the lines in the keyword file are read.
- The location of the found word or phrase that matches an entry in the keyword file is remembered to be used by subsequent actions.

Sample document for page identification with keywords

- The page can be identified by keywords.
- Example:
 - Use the text: **Rental Car Co.**

Thank you for renting with Rental E-Receipt Rental Agreement Number:703142974		Rental Car Co.	
Customer Information Customer Name: MILLARD BRYAN Work Address: 4650 GLUMACK DR Work Worksite Number: L193397 Customer Status: Rental Car co. FIRST Method of Payment: AMEX Credit Card #: CAXXXXXX0000XX1037 Free Traveler: DL02059946760		Vehicle Information Car Group Rented: FULL SIZE 4 DOOR Car Group Charged: FULL SIZE 4 DOOR Car Model: MINI TOW-A-PAIR 4DR Plate Number: MNV88CTE Car Number: 42457553 Mileage Out: 5109 Mileage In: 5263 Fuel Out: 8/8 Fuel In: 8/8 Mileage Driven: 174	
Rental Information Pickup Date/Time: 18APR102115 Pickup Location: MINNEAPOLIS-ST. PAUL INTL APO 4650 GLUMACK DR ST PAUL, MN 55111 US 912-726-3220		Return Date/Time: 23APR101551 Return Location: MINNEAPOLIS-ST. PAUL INTL APO 4650 GLUMACK DR ST PAUL, MN 55111 US 912-726-3220	
Vehicle Charges Time: 0 miles @ .00 0.00 TAX 7.275% 15.99 0 hours @ 18.01 0.00 #11.2% SURCHARGE 24.20 0 days @ 29.83 0.00 Fuel Service 0.00 1 weeks @ 179.00 179.00 Total Charges 260.05 Time & Mileage 179.00 \$ADY ERF 2.35 **10.00% FEE 18.51 LDW 0.00 FTF SRS 0.75DY 3.75 \$3.25 JOY CFC 16.25 SUBTOTAL 212.86 Amount Due 260.05			
<small>*CONCESSION RECOVERY FEE #12% VEH TAX & 5% REBATE FEE ENERGY RECOVERY FEE 47/DY CUST FEES 10.00 FF ML SPNTS EARNED 600</small> Car Rental E-Receipt <small>Please do not reply to this message. If you would like to receive an electronic version of this rental or if you do not wish to receive electronic receipts, please e-mail a.receipts@nix.com for assistance. If you have a question regarding this bill, please call us at 1-800-352-7800. This receipt reflects your charges as of the time of your return.</small>			

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-23. Sample document for page identification with keywords

The screen capture shows an example of a structured form with a Title (example for keywords) that can be used for page identification. Any other words on the page that are unique to this form can also be used to identify the page.

Page identification by using fingerprints

- Fingerprinting identifies the current page based on fingerprint match
- Datacap generates a fingerprint that describes each incoming page

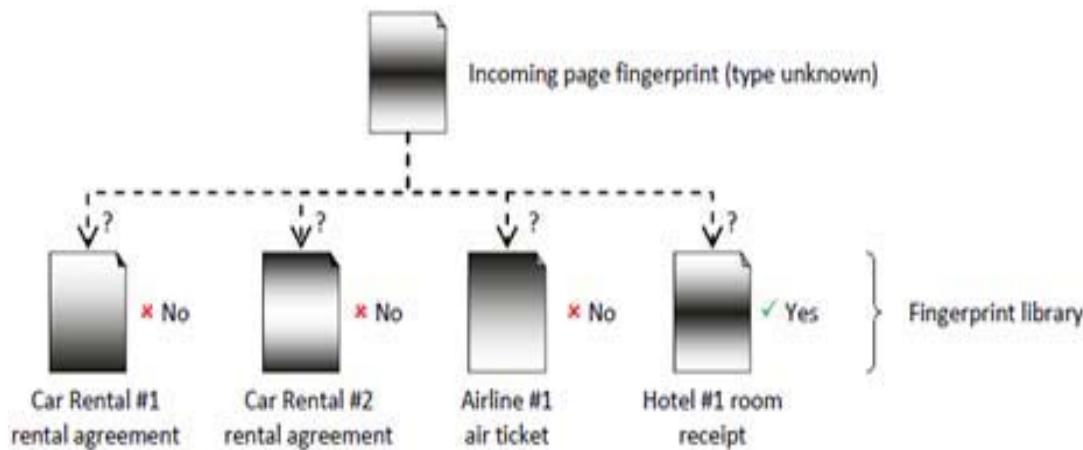


Figure 1-24. Page identification by using fingerprints

How does the Fingerprinting technique work?

The fingerprint can include information about the relative densities of different regions of the page or the location of text on the page.

Datacap compares an incoming page fingerprint to existing ones. If it matches an existing fingerprint, it is safe to assume that the incoming page is of the same class as the existing one. The offset required to give the best match is also captured. If a match does not occur, it creates a fingerprint.

Technique is adapted for structured and semi-structured documents with a fairly constant layout.

A fingerprint consists of an image file (.tif) and a recognition file (.cco)

- The CCO contains the location of all words and lines
- For example, compare dark and light zones to identify an incoming page

Example of page identification

In the example that is shown on the diagram, the incoming page matches the Hotel #1 room receipt. Datacap assigns it the type "Room_Receipt", and records the ID of the matching fingerprint in the runtime batch hierarchy.

Sample document for page identification with Fingerprints

- The page does not contain a barcode
- Can be identified by a fingerprinting

Thank you for renting with Rental E-Receipt		Rental Car Co.																																																																		
Rental Agreement Number: 703142974																																																																				
Customer Information Customer Name: MILLARD BRYAN Work Address: 4550 GLUMACK DR Work Worldwide Discount Number: L198397 Customer Status: Rental Car co. FIRST Method of Payment: AMEX Credit Card #: CAXXXXXX0000XX1037 Free Traveler: DL02059946760		Vehicle Information Car Group Rented: FULL SIZE 4 DOOR Car Group Charged: FULL SIZE 4 DOOR Car Model: MINI TOWER SAHR 4DR Plate Number: MNV88CTE Car Number: 42457553 Mileage Out: 5109 Mileage In: 5263 Fuel Out: 8/8 Fuel In: 8/8 Mileage Driven: 174																																																																		
Rental Information Pickup Date/Time: 18APR102115 Return Date/Time: 23APR101551 Pickup Location: MINNEAPOLIS-ST. PAUL INTL APO Return Location: MINNEAPOLIS-ST. PAUL INTL APO 4650 GLUMACK DR ST PAUL, MN 55111 US 612-726-3220																																																																				
Vehicle Charges <table> <thead> <tr> <th>Time:</th> <th>0 miles @ .00</th> <th>0.00</th> <th>TAX 7.275%</th> <th>15.99</th> </tr> </thead> <tbody> <tr> <td>0 hours @ 18.01</td> <td></td> <td>0.00</td> <td>#11.2% SURCHARGE</td> <td>24.20</td> </tr> <tr> <td>0 days @ 29.83</td> <td></td> <td>0.00</td> <td>Fuel Service</td> <td>0.00</td> </tr> <tr> <td>1 weeks @ 179.00</td> <td></td> <td>179.00</td> <td>Total Charges</td> <td>260.05</td> </tr> <tr> <td colspan="5">Time & Mileage 179.00</td> </tr> <tr> <td colspan="5">\$4.00Y ERF 2.35</td> </tr> <tr> <td colspan="5">**10.00% FEE 18.51</td> </tr> <tr> <td colspan="5">LDW 0.00</td> </tr> <tr> <td colspan="5">FTP SRS 0.75DY 3.75</td> </tr> <tr> <td colspan="5">\$3.25 JOY CFC 16.25</td> </tr> <tr> <td colspan="4">SUBTOTAL 212.86</td> <td>Amount Due 260.05</td> </tr> <tr> <td colspan="5">*CONCESSION RECOVERY FEE #2% VEH TAX & 5% REBATE FEE ENERGY RECOVERY FEE 47/DY CUST 10.00% FEE FF ML SPNTS EARNED 600</td> </tr> <tr> <td colspan="5">Car Rental E-Receipt <small>Please do not reply to this message. If you would like to receive an electronic receipt for this rental or if you do not wish to receive electronic receipts, please e-mail e.receipt@nix.com for assistance. If you have a question regarding this bill, please call us at 1-800-352-7800. This receipt reflects your charges as of the time of your return.</small> </td> </tr> </tbody> </table>				Time:	0 miles @ .00	0.00	TAX 7.275%	15.99	0 hours @ 18.01		0.00	#11.2% SURCHARGE	24.20	0 days @ 29.83		0.00	Fuel Service	0.00	1 weeks @ 179.00		179.00	Total Charges	260.05	Time & Mileage 179.00					\$4.00Y ERF 2.35					**10.00% FEE 18.51					LDW 0.00					FTP SRS 0.75DY 3.75					\$3.25 JOY CFC 16.25					SUBTOTAL 212.86				Amount Due 260.05	*CONCESSION RECOVERY FEE #2% VEH TAX & 5% REBATE FEE ENERGY RECOVERY FEE 47/DY CUST 10.00% FEE FF ML SPNTS EARNED 600					Car Rental E-Receipt <small>Please do not reply to this message. If you would like to receive an electronic receipt for this rental or if you do not wish to receive electronic receipts, please e-mail e.receipt@nix.com for assistance. If you have a question regarding this bill, please call us at 1-800-352-7800. This receipt reflects your charges as of the time of your return.</small>				
Time:	0 miles @ .00	0.00	TAX 7.275%	15.99																																																																
0 hours @ 18.01		0.00	#11.2% SURCHARGE	24.20																																																																
0 days @ 29.83		0.00	Fuel Service	0.00																																																																
1 weeks @ 179.00		179.00	Total Charges	260.05																																																																
Time & Mileage 179.00																																																																				
\$4.00Y ERF 2.35																																																																				
**10.00% FEE 18.51																																																																				
LDW 0.00																																																																				
FTP SRS 0.75DY 3.75																																																																				
\$3.25 JOY CFC 16.25																																																																				
SUBTOTAL 212.86				Amount Due 260.05																																																																
*CONCESSION RECOVERY FEE #2% VEH TAX & 5% REBATE FEE ENERGY RECOVERY FEE 47/DY CUST 10.00% FEE FF ML SPNTS EARNED 600																																																																				
Car Rental E-Receipt <small>Please do not reply to this message. If you would like to receive an electronic receipt for this rental or if you do not wish to receive electronic receipts, please e-mail e.receipt@nix.com for assistance. If you have a question regarding this bill, please call us at 1-800-352-7800. This receipt reflects your charges as of the time of your return.</small>																																																																				

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-25. Sample document for page identification with Fingerprints

The screen capture shows an example of a structured form without a barcode. A fingerprint can be used for page identification.

Automatic Recognition from Images

Barcode



OCR

Helen Sutton

ICR

First Name	L	O	R	I				
------------	---	---	---	---	--	--	--	--

Unconstrained

JOHN

Cursive

Madison

Natural
handwriting

Canyon Country

Figure 1-26. Automatic Recognition from Images

Automatic Classification reduces the workload of manual processing.

OCR – Optical Character Recognition

- This technology is used to convert system-printed text in an image to editable text
- There are several OCR engines: OCR_a, OCR_sr, and OCR_j
- There are two types of engines for OCR_a and OCR_sr: Insight Edition (segment-based), and standard OCR edition (line-based)

ICR – Intelligent Character Recognition

- Recognize hand written characters

OMR – Optical Mark Recognition

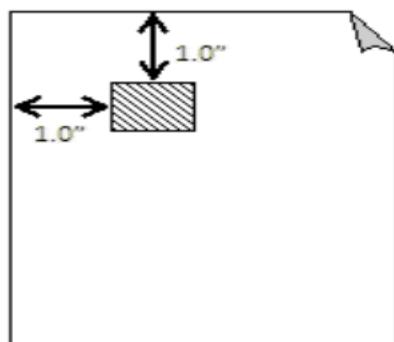
- Recognize checkboxes and radio buttons.

Barcode recognition

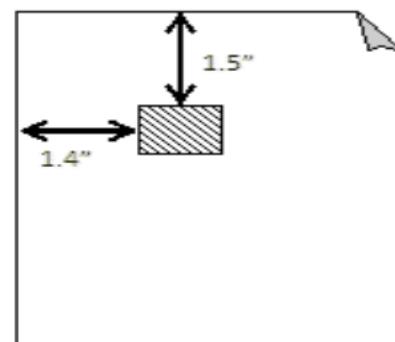
- The ability to recognize and interpret barcodes.

Page identification with pattern recognition

- Pattern recognition identifies the current page based on geometric patterns such as:
 - Vendor logos
 - Page registration marks
 - Text-based patterns



Fingerprint



Scanned page

Figure 1-27. Page identification with pattern recognition

You can use Datacap pattern matching to identify pages and adjust misaligned or distorted images.

Page identification with separator sheets

- Identifiable separator sheets are physically placed in the batch so that they can be used to identify the following page.

Page identification with manual identification

- Final alternative might be identification by a human.

Figure 1-28. Page identification with separator sheets

Extraction

- Recognize zones
- Locate action library
 - Per field by using text analytics
- Clean and filter
 - Clean
 - Filter
- Validate
- Route

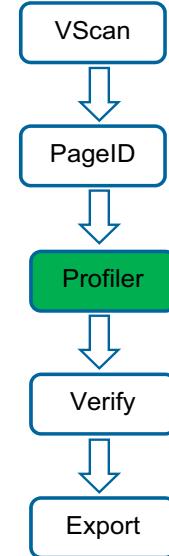


Figure 1-29. Extraction

Recognize zones

Use fingerprints, the FPXML actions that are better for memory, and the use FPXML setting in Application Manager.

Locate action library

The actions use text analytics. You locate keys and values for each field.

Clean and filter

Clean removes unwanted characters from a data field value. For example, it removes "\$" from the cost value. Filter ensures that the correct format is retained. For example, if there are any extra spaces or characters, filter removes it.

Configure the rules for clean and filter so that they are easily adaptable. for example, Clean if you need to change O with zero or I to one.

Validate

After clean and filter, the data is validated to ensure it is the correct data. For example, URL or Phone number.

Route

After validation, you can route to verify. If there are multiple errors, you can route to Fixup task.

Data verification

- Based on the validation results, you decide to route the data to an operator for manual verification
- In the verification step, the operator corrects and validates the fields
- You can have more than one verification step
- You can customize the verification (panels) interface

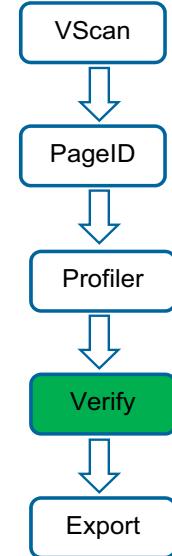


Figure 1-30. Data verification

You can do the manual verification of the documents either in the Datacap Navigator (web client) or in the Datacap Desktop (thick) clients.

By configuring queue by user, you can get the data verified by two or more different users.



Data export

- Captured Documents and Data can be exported to repositories



- Captured Documents and Data can be used in other applications

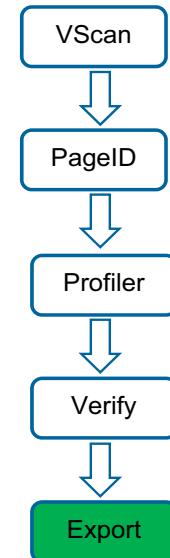


Figure 1-31. Data export

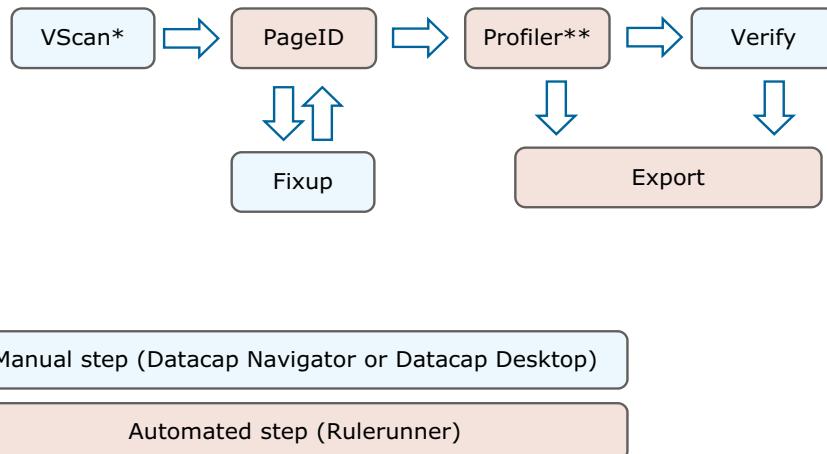
Captured documents and data can be exported to repositories such as:

- IBM FileNet Content Manager
- IBM Content Manager
- IBM Content Manager OnDemand (CMOD)
- Contents on BOX
- Microsoft SharePoint

Captured documents and data can be used in other applications such as:

- SAP
- Case-based and Business Process Management solutions
- Relational databases
- XML export
- Electronic Data Interchange (EDI)
- LOB applications such as SAP, Account Servicing, Policy Administration, Insurance Underwriting, Claims, Loan Origination, and Benefits

Manual and automated steps in Datacap batch



* - The step can be automated

** - Extract step is also known as Profiler

Figure 1-32. Manual and automated steps in Datacap batch

The diagram shows the workflow steps for a Datacap application.

- The manual steps such as VScan and Verify are indicated in blue.
- The automatic steps such as Page ID, Profiler (also known as Profiler), and Export are indicated in orange. The automatic steps are run by Rulerunner.
- In addition, the Fixup task is included. If there is low level of confidence, the Page ID step can be routed to the Fixup task.
- You can also add more verification steps. For example, queue by user filter. That allows two different users to verify the data.
- If the confidence level is high, the workflow can skip the manual verification step and go to Export step directly. It depends on how well the document is structured, how good is OCR quality, and how well the label-value pairs are defined.



Note

The steps that are presented in this slide are high-level generic steps. You can customize them by adding more steps and renaming the steps to suit your business needs.

1.3. Datacap architecture



Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-33. Datacap architecture

Single system configuration

- Datacap system needs two main services:
 - Datacap service (main)
 - Rulerunner service
(to run background tasks)
- User Environment
 - Datacap Desktop (thick client)
 - Datacap Navigator (thin client) that requires:
 - Web Application Server
 - wTM (REST service)



Figure 1-34. Single system configuration

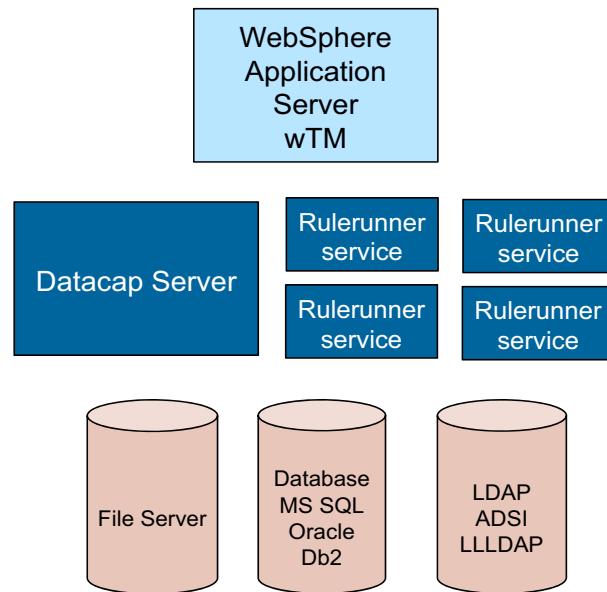
In a Single system configuration, all the components are installed on the same system.

The configuration is used for providing product demonstrations, in a proof-of-concept environment, during initial product evaluation, or development environment.

In this configuration, the Rulerunner can run only two threads and user environment can have only few users.

Multiple-system configuration

- Each Datacap component in separate systems:
 - Datacap service (main)
 - Rulerunner services (to run background tasks)
- User Environment
 - Datacap Navigator (thin client) that requires:
 - Web Application Server
 - wTM (REST service)



[Introducing IBM Datacap](#)

© Copyright IBM Corporation 2020

Figure 1-35. Multiple-system configuration

In a multiple system configuration, the components are installed on dedicated systems.

- For a production system, always host the Datacap server and Rulerunner services in separate systems.
- In this configuration, you can have many systems for Rulerunner servers to handle many threads.
- It supports for hundreds of simultaneous users.

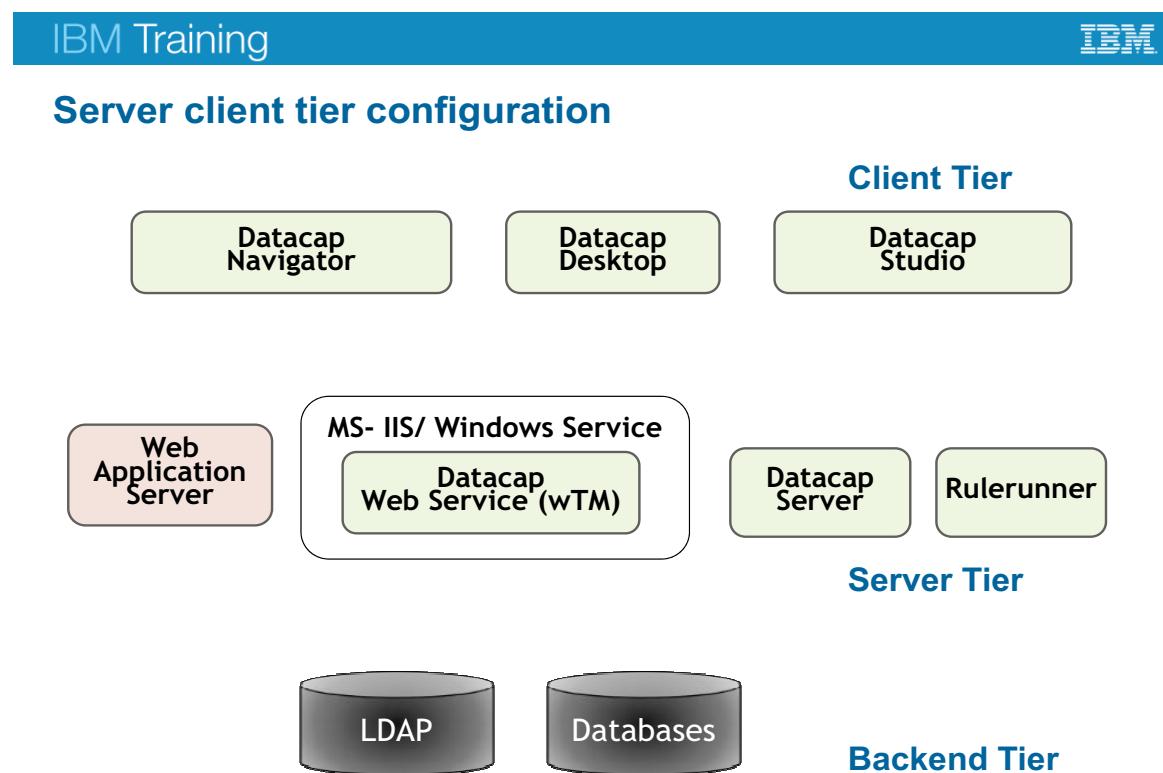


Figure 1-36. Server client tier configuration

The screen capture shows the server client tiers for a Datacap system.

Client Tier

Datacap Navigator and Datacap Desktop are user environments for processing batches.
Datacap studio is development environment for creating Datacap applications.

Server Tier

- Datacap Server is the main server for all Datacap components. It does the authentication, batch task-queuing, database access, and file sharing.
- Rulerunner is the Datacap background processing engine. You must configure Rulerunner to be able to run background tasks without user intervention.
- Web Application Server and Datacap Web Service (wTM) are required for Datacap Navigator.

You can host wTM as a Windows service or wTM can be run on Microsoft Internet Information Server (IIS). wTM is required for web and mobile client connection. RESTful API services for external access to Datacap services and also used by IBM Content Navigator to provide Datacap Navigator and Mobile connectivity.

Backend Tier: Database backend is required for all the components. File server and other repository servers are also used.

Components – Datacap Web Services

- Datacap Web Services (wTM)
- Windows web service or Microsoft IIS-based web service
- For interaction with Datacap through a simple REST API
 - Working with batches
 - Editing the workflow
- Datacap Web Services support HTTP and HTTPS protocols

Figure 1-37. Components – Datacap Web Services

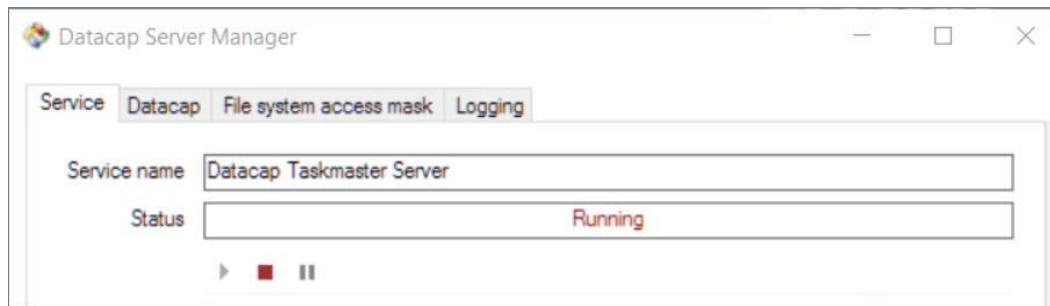
Datacap Web Services are also called wTM.

It is a Windows web service or Microsoft IIS-based web service for interaction with Datacap through a simple, platform-independent, representational state transfer (REST) application programming interface (API)

Datacap Navigator and custom applications use RESTful web service interface.

Components – Datacap Server

- Controls authentication, database access, and batch-queuing
- Accesses the administration and engine databases
- Accesses the file system for thin clients
- Runs as a Windows service



[Introducing IBM Datacap](#)

© Copyright IBM Corporation 2020

Figure 1-38. Components – Datacap Server

Datacap Server:

Accesses the administration and engine databases and all other components must access the databases through Datacap Server.

Accesses the file system for thin clients. Datacap Navigator access application and document files through the Datacap server

The main application settings file datacap.xml is on the file share where the application files are stored. In smaller systems, the file is on a Datacap Server. In medium or large systems, the file is on a dedicated file server or a NAS or SAN device.

In the Datacap Server Manager, you can start, pause, and stop the Datacap Server service. You can set the System event log and the Datacap log.

In the Advanced settings pane of the Datacap Server Manager, you can configure the Datacap Server service connections port and set the database command timeout. You can also select the authentication method and configure the Batch naming template.



Components – Rulerunner Server

- Runs batch processing tasks that do not require operator interaction
 - Examples: Page ID, Profiler (also known as extract), and export
- Runs as a Windows service

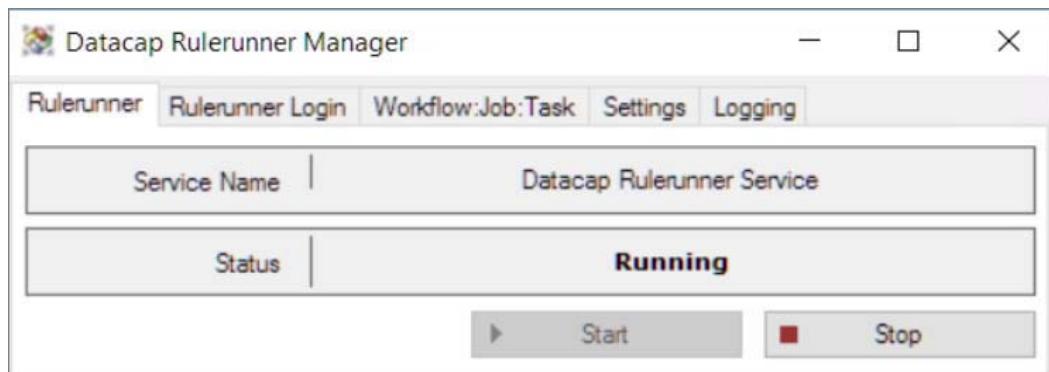


Figure 1-39. Components – Rulerunner Server

Rulerunner is a unified Windows service that runs background tasks that do not require operator interaction.

On the Rulerunner Manager, you can:

- Start and stop the Rulerunner Service
- Decide which applications, workflows, and tasks are processed by the Rulerunner Service
- Configure settings for how the Rulerunner Service processes tasks
- Set Rulerunner logging

The Datacap.xml file

- This file is the directory of all the applications in a Datacap environment.
- It contains:
 - The Datacap version number
 - The name and folder location of each application in the environment.

```

1  <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2  <datacap ver="9.0">
3      <app name="FieldTest" ref="FieldTest"></app>
4      <app name="MyFormApp" ref="C:\Datacap\MyFormApp"></app>
5      <app name="MyCMISApp" ref="C:\Datacap\MyCMISApp"></app>
6      <app name="Flex" ref="Flex"></app>
7      <app name="Forms" ref="Forms"></app>
8      <app name="Learning" ref="Learning"></app>
9      <app name="TravelDocs" ref="TravelDocs"></app>
10     <app name="APT" ref="APT"></app>
11     <app name="Medical Claims" ref="Medical Claims"></app>
12     <app name="FormTemplate" ref="Templates\FormTemplate"></app>
13     <app name="LearningTemplate" ref="Templates\LearningTemplate"></app>
14 </datacap>

```

Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-40. The Datacap.xml file

The screen capture shows a sample datacap.xml file.

- The datacap.xml file is in the Datacap root folder. It contains the Datacap version number and a tag that defines the name for each of the defined Datacap Capture applications.
- The version and exact build number are also shown in the C:\Datacap\version.txt file.
- The ref attribute of the app node in datacap.xml uses a relative path if no folder is specified.



Important

The Datacap Application Service is case-sensitive. When you add or change entries in the datacap.xml file, make sure that the case matches the case of the UNC paths, folders, and file names.

Unit summary

- Describe capture concepts and Datacap
- Identify the tasks in Datacap process
- Understand Datacap Architecture
- Identify Datacap service components

Figure 1-41. Unit summary

Review questions

1. True or False: Datacap supports both structured and unstructured documents.
2. Which of the following items are Datacap page identification methods? (Choose three)
 - A. Keyword
 - B. Batch Process
 - C. Pattern recognition
 - D. Fingerprint
 - E. Document export
 - F. Input channel
3. True or False: Datacap captured documents and data can be exported to content repositories or can be used in other applications.



Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-42. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. True or False: Datacap supports both structured and unstructured documents.
The answer is True.

2. Which of the following items are Datacap page identification methods? (Choose three)
 - A. Keyword
 - B. Batch Process
 - C. Pattern recognition
 - D. Fingerprint
 - E. Document export
 - F. Input channel**The answer is A,C, and D.**

3. True or False: Datacap captured documents and data can be exported to content repositories or can be used in other applications.
The answer is True.



Introducing IBM Datacap

© Copyright IBM Corporation 2020

Figure 1-43. Review answers

Unit 2. Introducing Datacap clients for processing batches

Estimated time

00:30

Overview

This unit introduces the Datacap clients for processing batches.

How you will check your progress

- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Identify the functions of Datacap Desktop client
- Identify the functions of Datacap Navigator client

Figure 2-1. Unit objectives

Topics

- Datacap Desktop
- Datacap Navigator

Figure 2-2. Topics

2.1. Datacap Desktop

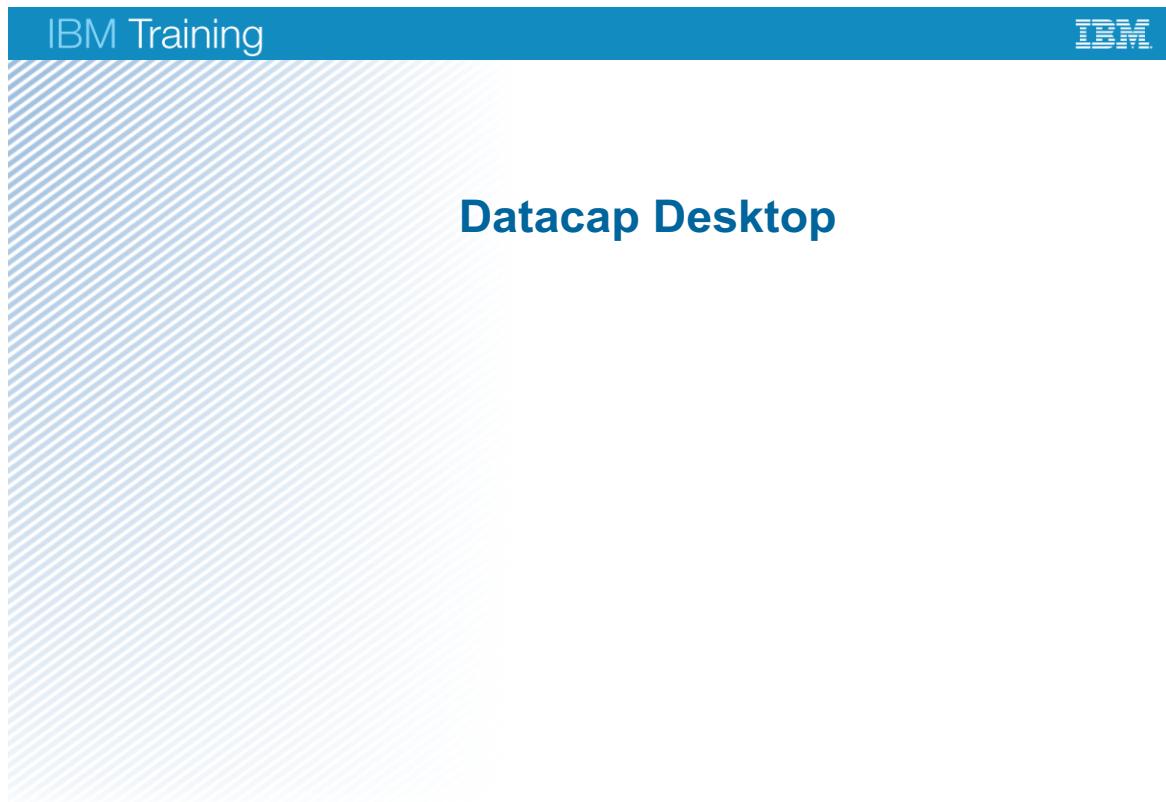


Figure 2-3. Datacap Desktop

Datacap Desktop capabilities overview

- Use Datacap Desktop client to do the following tasks:
Create batches, identify pages, and export data and documents
- To use Datacap Desktop to complete a task:
Must specify Desktop as the client in setup page for that task
- The Datacap Desktop:
 - Has a single login point for completing tasks
 - Can select any application that is configured in the Datacap system for processing tasks
 - Includes the Queue Monitor in which users with appropriate privileges can view or start the pending tasks for the selected application

Figure 2-4. Datacap Desktop capabilities overview

Datacap Desktop is one of the clients that you can use to process the batches for the Datacap applications. It is a thick client.

Use Datacap Desktop client to do the following tasks:

- Create batches
It is the process of scanning batches of paper documents or accessing a folder of electronic documents and queuing them for further processing.
- Identify pages
Identify a page by the layout or identifiable fields so that further processing can be done to extract and verify the data on the page.
- Create Documents
Group pages into documents and create a document that represents the grouped pages.
- Export data and documents
Batches of processed documents are sent on the next step in the business process. The next step might be to export the documents to a repository where they are stored and can be searched, sorted, and categorized for ease of access. The next step might also be direct them to subsequent application for further processing.

Use Datacap Desktop with your Datacap applications

Use Datacap Desktop to complete the following tasks:

- Vscan
 - Imports files from a file system
- Scan
 - Imports physical paper documents by scanning
- PageID
 - Identifies the page type of a scanned image
- Profiler
 - Creates documents, extracts and validates data and routes documents to the next step
- Verify
 - A user can verify, and correct document and extracted data values
- Export
 - Exports batches to a specified location

Figure 2-5. Use Datacap Desktop with your Datacap applications

VScan

This task is for a virtual scan that imports files from a specified directory.

Scan

In production environments, a scan task is configured to scan paper documents that use scanners.

Profiler

- Locates and extracts data values from pages
- Arranges identified pages into documents
- Does confidence testing on all characters read from the page
- Validates the extracted field values
- Routes documents to the next task based on validation results (optional step)

The documents that are error-free can be routed directly to the Export task. The documents that have errors are routed to an operator to verify and correct potential error.

Verify

- This task requires user input to correct any errors or integrity issues that a preceding task encounters.
- The batch does not continue to the next task until the Fixup task is completed.

Background

- The task completes all tasks that do not require user intervention. For example, PageID, Profiler, and Export.
- It automates the completion of pending batches.

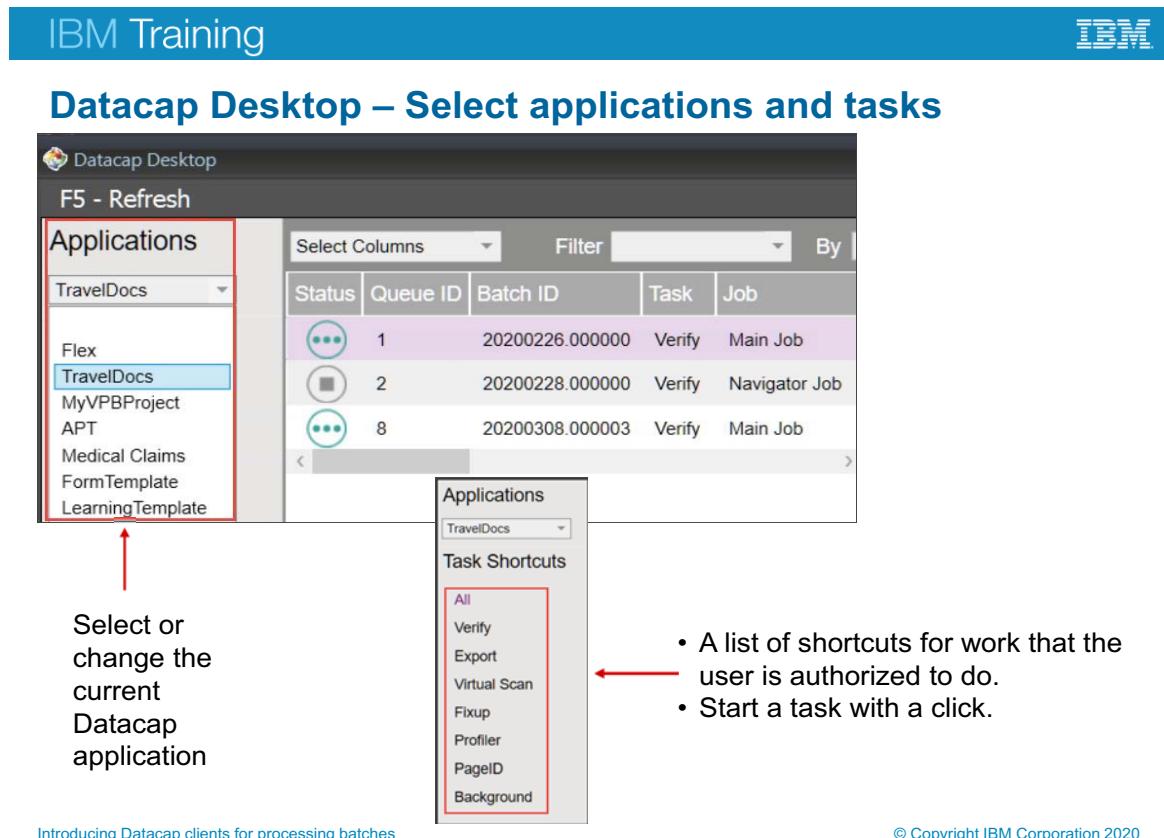


Figure 2-6. Datacap Desktop – Select applications and tasks

The upper screen capture shows the selection of an application in the Datacap Desktop client and the lower screen capture shows Tasks shortcuts.

Desktop allows the users to switch between applications and tasks with ease. Users can select an application or change to another application from the list. When you select an application, a list of shortcuts for work that the user is authorized to do for the application is displayed.

When the user clicks a shortcut to start, the batch items are displayed in the view on the right pane. The list depends on the user permissions, application, Task Shortcuts selections, and the mode (manual, automatic, or manual for hold) of the task.

Job Monitor view

- If user has Job Monitor view privilege, the "All" shortcut is listed.
- When the user selects "All", the Job Monitor view is shown in the right pane.
- The view lists the batches for all the allowed job-task combinations for the user.

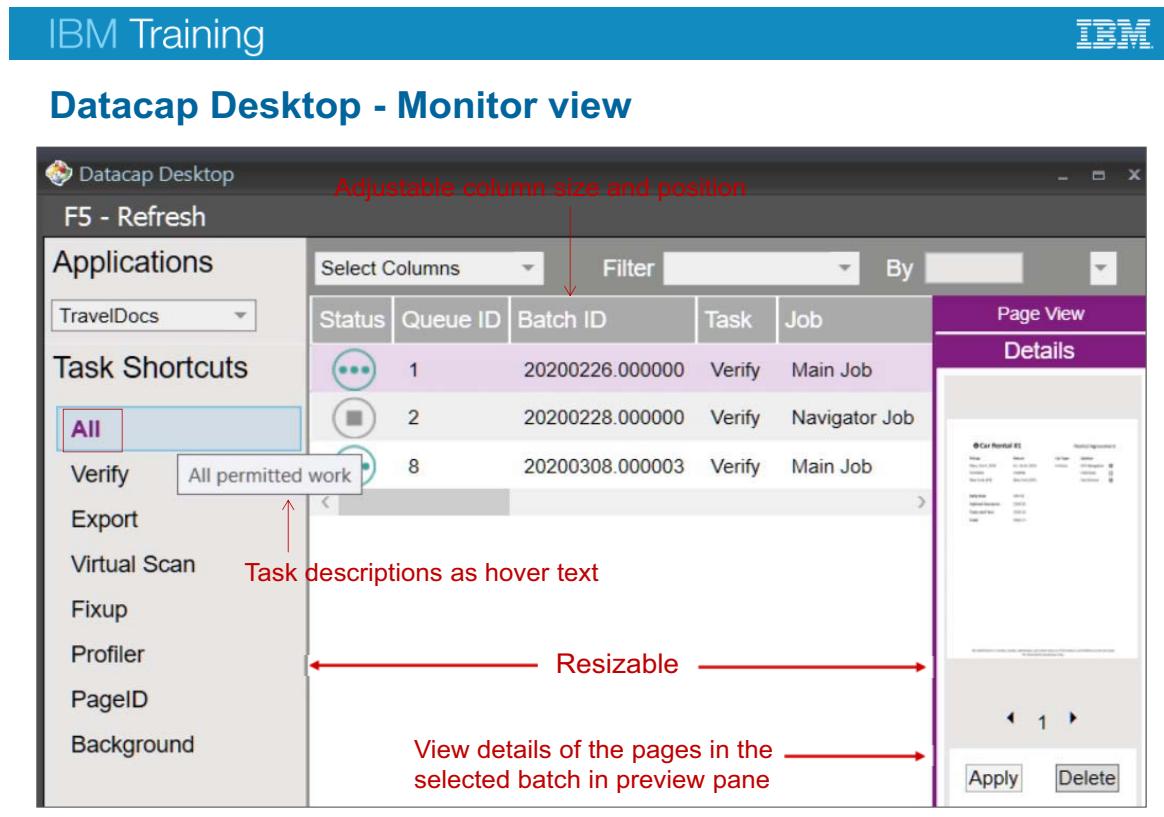


Figure 2-7. Datacap Desktop - Monitor view

The screen capture shows the Datacap Desktop client.

Batches in Monitor view

The status of a batch is shown as an icon in the first column. In the list, the pending task has a different color that is compared to the task on hold. When you hover over a task, description for that task is shown. Double-click of a batch starts the associated task.

Task Preview

- Single click of a batch shows its details in the preview pane on the right.
- The preview has two tabs:
 - Details View shows the batch details such as the number of pages and documents.
 - On the Page View, you can look at the pages in a batch without having to start a task.
- Preview pane is not shown when no batch is selected. You can resize and collapse the preview pane.

Column size and position

- Drag the individual column by header to change its relative position in the table.
- Click a column header to sort by that column.

IBM Training IBM

Select columns and apply filter in the Monitor view

The screenshot shows the Datacap Desktop client interface. At the top left is a 'Task Shortcuts' sidebar with options like All, Verify, Export, Virtual Scan, Fixup, Profiler, and PageID. To its right is a 'Select Columns' dialog box containing a list of items with checkboxes. Below these are three red annotations: 'Apply filter' with a downward arrow pointing to the 'Filter' dropdown; 'Specify count' with a downward arrow pointing to the 'Items per page' dropdown set to 15; and 'Page Navigator' with a downward arrow pointing to the page navigation controls. The main area displays a table titled 'Monitor' with columns for Status, Queue ID, Batch ID, Task, Job, Job Start, Job Time, Task Start, Operator, Station, and Doc. The table contains three rows of data.

Status	Queue ID	Batch ID	Task	Job	Job Start	Job Time	Task Start	Operator	Station	Doc
	8	20200308.000003	Verify	Main Job	3/8/2020 8:34 AM	177	3/8/2020 8:38 AM			1
	2	20200228.000000	Verify	Navigator Job	2/28/2020 3:06 AM	569	3/10/2020 7:42 AM	admin	1	1
	1	20200226.000000	Verify	Main Job	2/26/2020 10:01 PM	100	2/27/2020 7:52 PM			1

Figure 2-8. Select columns and apply filter in the Monitor view

The upper screen capture shows the selection of the columns to display in the Datacap Desktop client and the lower screen capture shows the filter and page counts.

2.2. Datacap Navigator

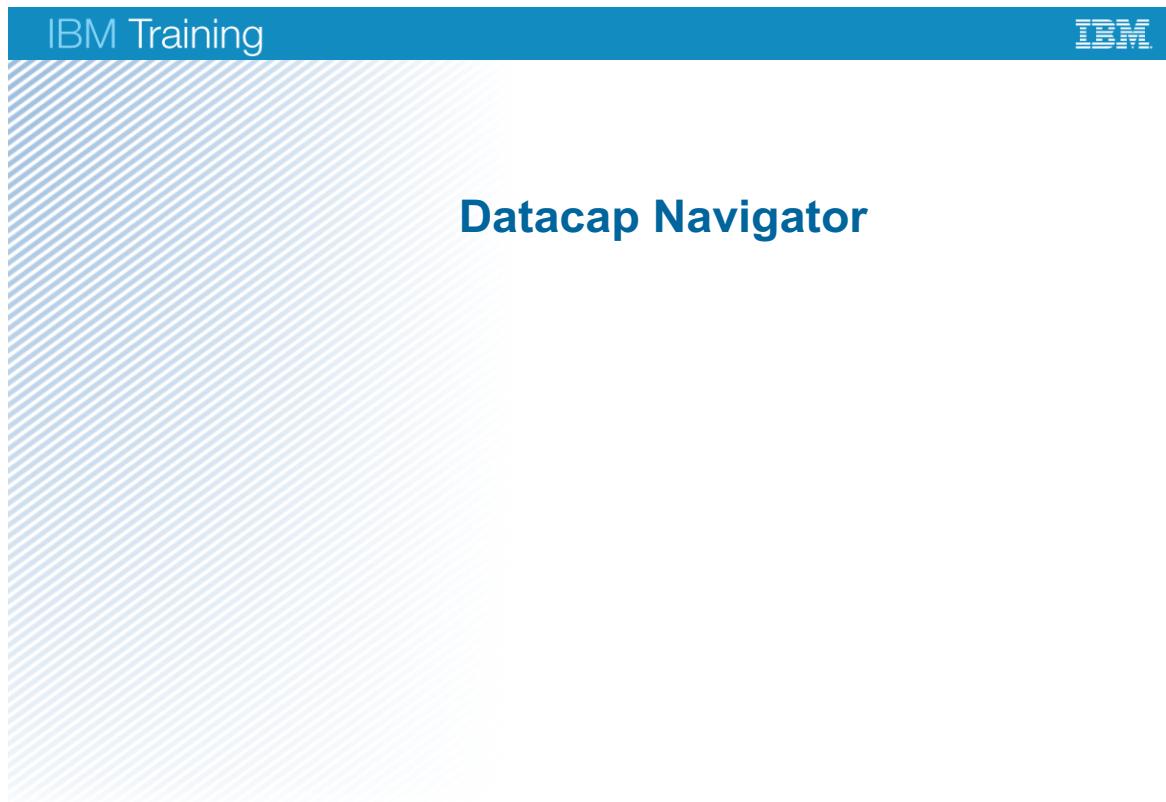


Figure 2-9. Datacap Navigator

What is Datacap Navigator?

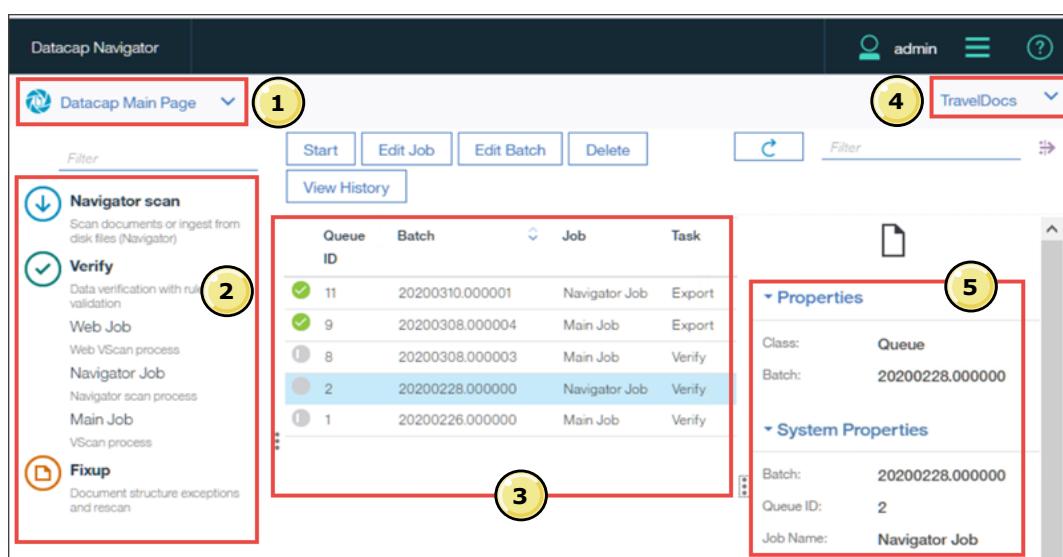
- IBM Content Navigator is a web-based client and framework.
To work with content from content management repositories
- Datacap includes a plug-in that operates within IBM Content Navigator.
- Datacap Navigator provides a web user interface that:
 - Supports both user and administrative features for Datacap
 - Provides access to the Datacap Job Monitor and to run tasks
 - Enables you to add users and groups
 - Provides access to configure stations and workflows

Figure 2-10. What is Datacap Navigator?

IBM Training



Datacap View for the business users



Queue ID	Batch	Job	Task
11	20200310.000001	Navigator Job	Export
9	20200308.000004	Main Job	Export
8	20200308.000003	Main Job	Verify
2	20200228.000000	Navigator Job	Verify
1	20200226.000000	Main Job	Verify

Introducing Datacap clients for processing batches

© Copyright IBM Corporation 2020

Figure 2-11. Datacap View for the business users

The Datacap view that is shown in the screen capture is for Datacap business users.

1. Feature List

- Datacap Navigator by default shows the Datacap Main Page.
- You can optionally add the Browse and Search pages that are part of IBM Content Navigator
- The Browse and Search pages helps to view the documents that are exported from Datacap process to repositories

2. Shortcut pane

- It contains a list of shortcuts for operations
- The pane shows a list of all of the tasks that they are authorized to run

3. Job Monitor

The list of the batches that are in process and the completed ones.

4. Select Application

- The list of Datacap applications that are available are listed
- Select an application and Job Monitor lists the batches for that application
- You can monitor jobs for the batches that are processed in Datacap Desktop

5. Properties pane

When you select a batch from the Job Monitor, you can view the details about the batch.

Views in Datacap Navigator

- In Datacap Navigator, there are two views (feature)
 - Datacap view for the business users
 - Datacap Administration view for administrators
- You can configure and customize Datacap Navigator for different roles
 - Business Users (To scan, verify, and monitor jobs)
 - Administrators (To configure and administer workflows for the application)
- You can add more features such as Browse or Search
 - To browse folders or search for documents that are exported
 - To scan the documents directly from the IBM Content Navigator

Figure 2-12. Views in Datacap Navigator

Completed Datacap documents can be exported to a repository for storage. These repositories can be accessed in IBM Content Navigator.

The Content Management capabilities in IBM Content Navigator open the Datacap documents in Search or Browse Views.

The screenshot shows the Job Monitor interface in Datacap Navigator. The toolbar at the top includes buttons for Start, Edit Job, Edit Batch, Delete, and View History. The job list table has columns for Queue ID, Batch, Job, Task, Status, and Job Start. The first two rows are highlighted with a red border. The right pane displays job properties (Class: Queue, Batch: 20200228.000000) and system properties.

Queue ID	Batch	Job	Task	Status	Job Start
11	20200310.000001	Navigator Job	Export	Job done	3/10/2020, 4:25 AM
9	20200308.000004	Main Job	Export	Job done	3/8/2020, 8:51 AM
8	20200308.000003	Main Job	Verify	pending	3/8/2020, 8:34 AM
2	20200228.000000	Navigator Job	Verify	hold	2/28/2020, 2:06 AM
1	20200226.000000	Main Job	Verify	pending	2/26/2020, 9:01 PM

Properties:

- Class: Queue
- Batch: 20200228.000000

System Properties:

Figure 2-13. Job Monitor

The screen capture shows the Job Monitor in Datacap Navigator.

1. Toolbar

You can do the following tasks: Start a job, Edit Job, Edit Batch, Delete Batches, View History

2. Job list

It lists all or filtered jobs.

3. Refresh

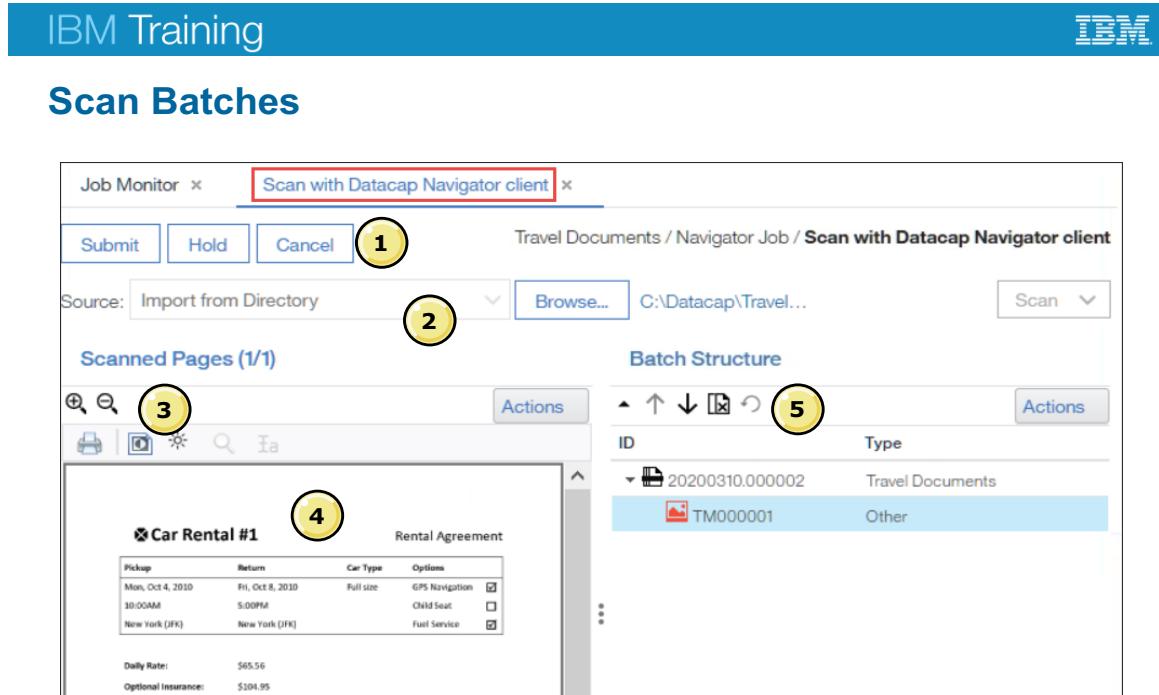
You can refresh the job list.

4. Filter

You can filter the batch items with quick search.

5. Job information pane

It shows job properties



Introducing Datacap clients for processing batches

© Copyright IBM Corporation 2020

Figure 2-14. Scan Batches

The screen capture shows the Scan page in Datacap Navigator.

1. Toolbar

You can submit, hold, or cancel a batch.

2. Scanner selector

It lists all available scanners. You can browse to local directory to get stored images.

3. Viewer toolbar

Zoom In and Zoom Out, Fit to Width & Height, Previous & Next Page

4. Image Viewer

View the scanned page in the viewer

5. Batch structure toolbar

Delete one or all pages, Move Up, Move Down to reorder the pages

IBM Training

Rescan options in the Scan tab

TravelDocs

Job Monitor x Scan with Datacap Navigator client x

Submit Hold Cancel Travel Documents / Navigator Job / Scan with Datacap Navigator client

Source: Import from Directory Browse... C:\Datacap...

Scanned Pages (1/1)

Scan Append Insert Replace

Batch Structure

ID	Type
20200310.000000	Travel Documents
TM000001	Other

Scanned Pages (2/2)

Actions

Introducing Datacap clients for processing batches

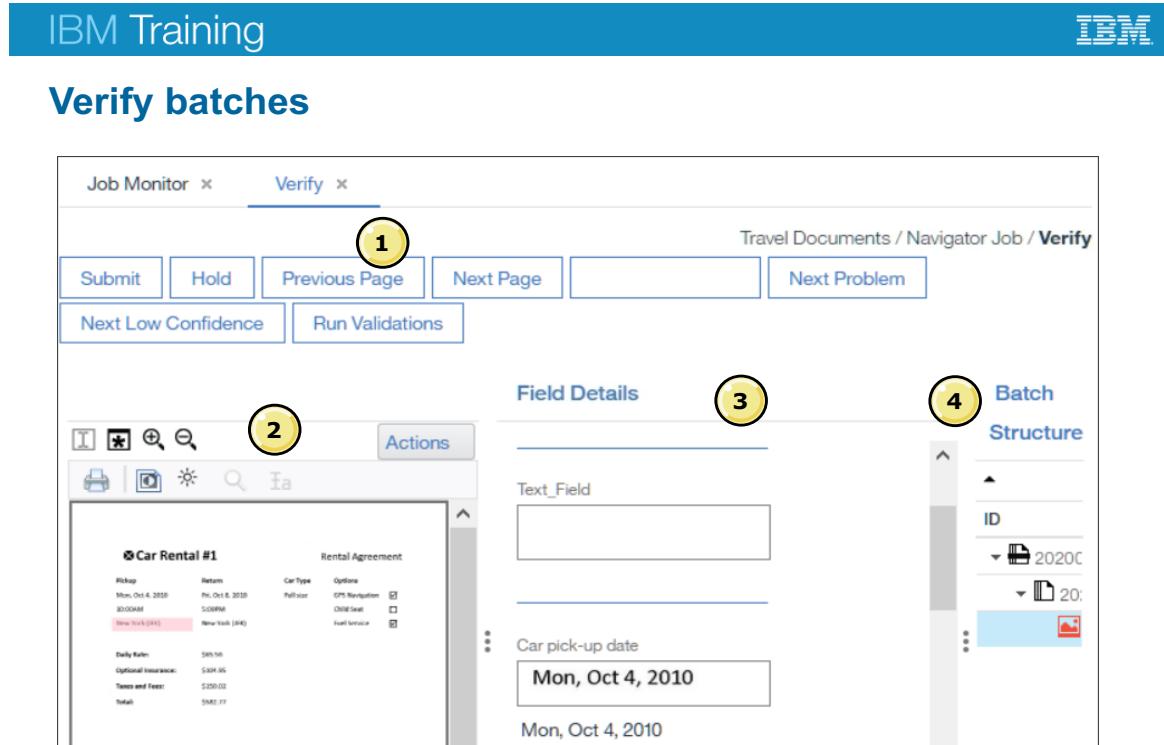
© Copyright IBM Corporation 2020

Figure 2-15. Rescan options in the Scan tab

The screen capture shows the scan page in Datacap Navigator client. After you scan the first image, you can scan again either to append or insert as the second image.

When you add the second image, the Scanned Pages label shows the count as 2/2 as shown in the second screen capture.

You can also scan again and replace the existing image.



Introducing Datacap clients for processing batches

© Copyright IBM Corporation 2020

Figure 2-16. Verify batches

The screen capture shows the Verify page in Datacap Navigator client.

It contains the following areas:

1. Toolbar

Submit batch, Hold batch, Previous & Next Page, Previous & Next Problem, Next Low Confidence Field, Run Validation on current page

2. Image Viewer

- In the viewer, the area of the field that is selected in the Field Details panel is highlighted.
- In the viewer, update the value for a field with clicking or drawing a rectangle on some words. This method of identifying a field is also referred to as the click'n'key method.
- In the toolbar, you can use various actions: Zoom In, Zoom Out, Fit to Width and Height, go to Previous & Next Page, Show all fields, Show recognized words, Show recognized lines, Show viewer in dual monitor

3. Field Details pane

- Shows the values for the fields
- Low confidence fields have a yellow background
- The fields that fail validation have a light pink background:

4. Batch structure pane

The document and pages are listed for the batch.

The screenshot shows the 'User settings' page in the Datacap Navigator client. At the top left is the 'IBM Training' logo, and at the top right is the 'IBM' logo. The main title is 'User settings'. On the left, a sidebar menu includes 'Change Password', 'Change Language and Locale Settings', 'Change User Settings' (which is highlighted with a red border), and 'Log Out'. The main content area is titled 'Settings' and shows the user 'admin' and repository 'TravelDocs'. It has tabs for 'Global', 'Scan', 'Upload', 'Classify', 'Verify', and 'Batch Prep'. Under 'Global', there is a section for 'Temporary directory for scanned images' set to 'c:\datacap\scan'. There are three checkboxes: 'Hide the title bar in all widgets' (unchecked), 'Hide the Shortcuts pane when you start a task' (unchecked), and 'Specify field values by using data type controllers' (checked). Below this is a section for 'Configure station IDs for applications' with a note about logging in again after changes. A dropdown menu for 'Station ID' shows the value '1'.

Figure 2-17. User settings

The screen capture shows the User Settings page in Datacap Navigator client.

Each user can change the settings to tailor the experience to meet the individual needs and to change the appearance and operation of the user interface.

Some of the common options that you can configure in User Settings are:

- Hide shortcuts pane when you start a task
- Set the number of pages to scan for a batch
- What columns are available in the Job Monitor

Unit summary

- Identify the functions of Datacap Desktop client
- Identify the functions of Datacap Navigator client

Figure 2-18. Unit summary

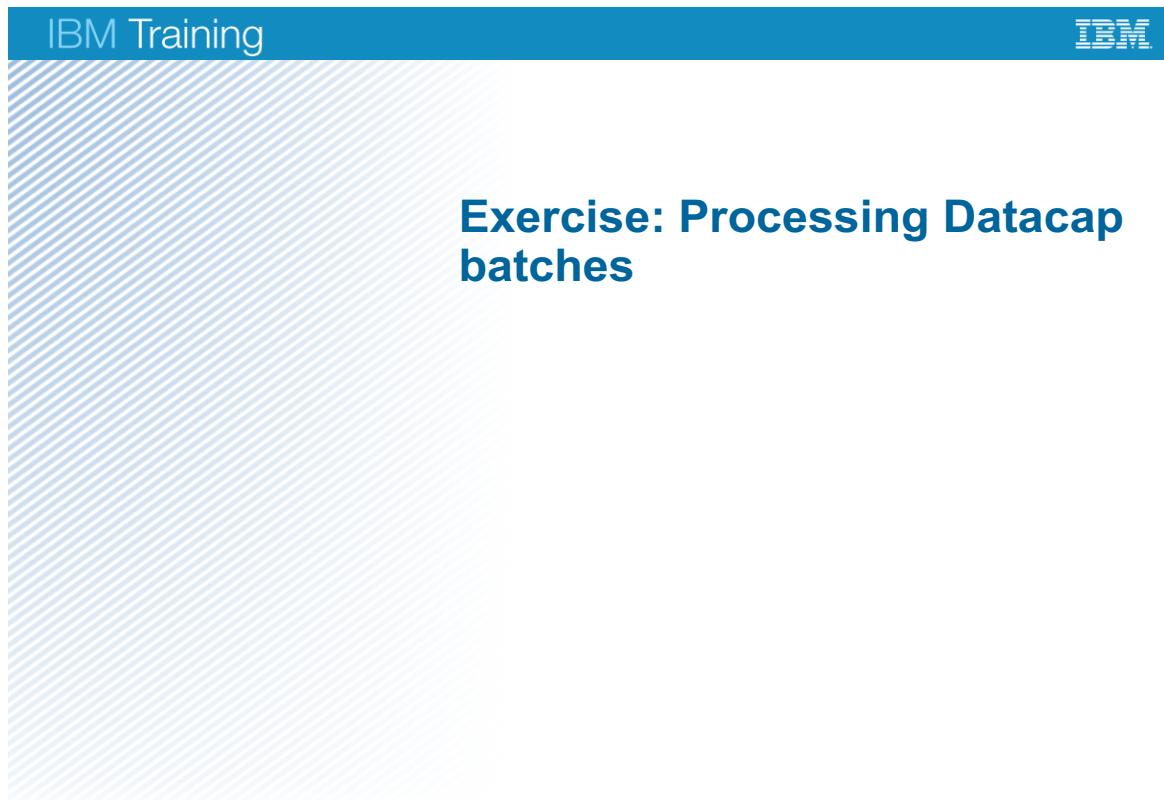


Figure 2-19. Exercise: Processing Datacap batches

Exercise objectives

- Process a batch in Datacap Desktop
- Process a batch in Datacap Navigator



Figure 2-20. Exercise objectives

In this exercise, you use the Datacap clients Datacap Desktop and Datacap Navigator to process the batches.

Scenario for the TravelDocs application

Assume that a company has numerous external sales, and technical support personnel who are continually traveling between customer sites. So, they have many car rental documents, airline ticket receipts, and hotel receipts from multiple travel companies.

- A Datacap application is already created that recognizes car rental agreements and air ticket receipts for the preferred vendors.
- The application is able to recognize rental details and total cost of service charged.
- The exported data from the scanned image is stored as text files on a file system.

Unit 3. Introducing Datacap Studio and application design

Estimated time

01:30

Overview

This unit introduces the development environment Datacap Studio and describes the Datacap application design principles.

How you will check your progress

- Review

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Identify the application development features in Datacap Studio
- Describe the Datacap application design concepts

Figure 3-1. Unit objectives

Topics

- Introduction to Datacap Studio
- Datacap Studio - Rulemanager
- Datacap Studio - Zones
- Datacap application design

Figure 3-2. Topics

3.1. Introduction to Datacap Studio

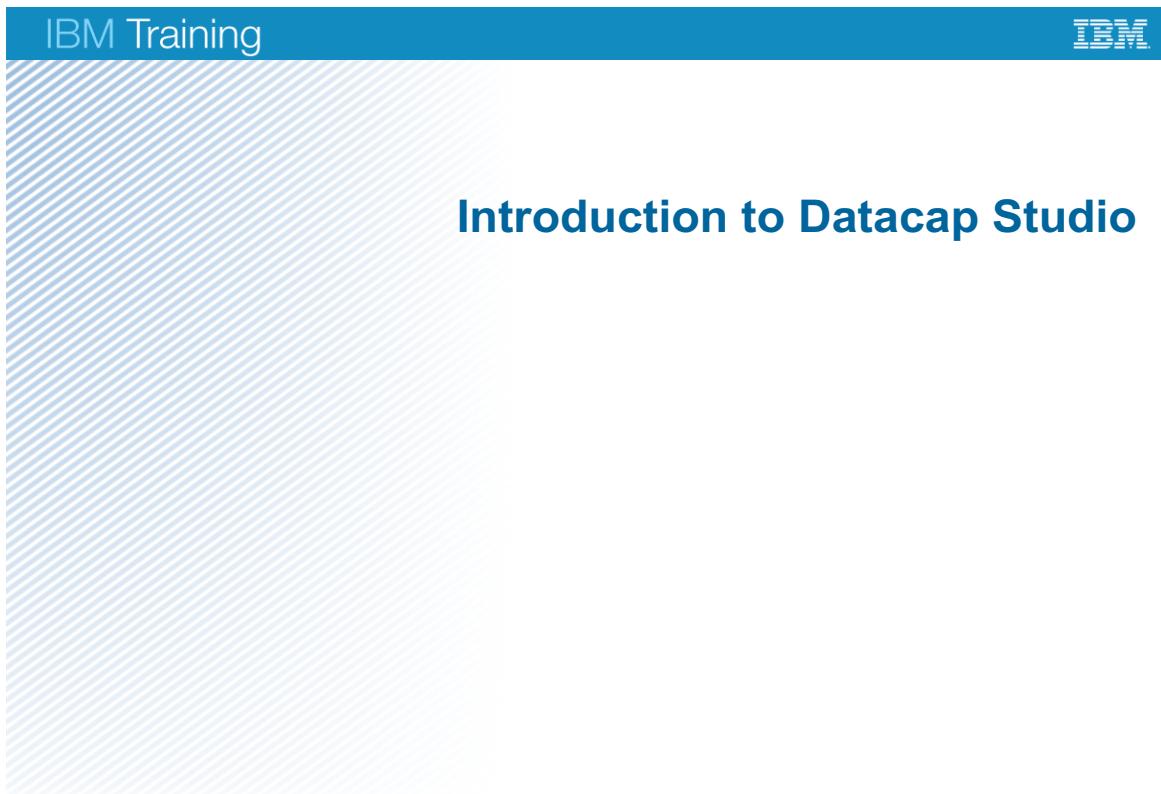


Figure 3-3. Introduction to Datacap Studio

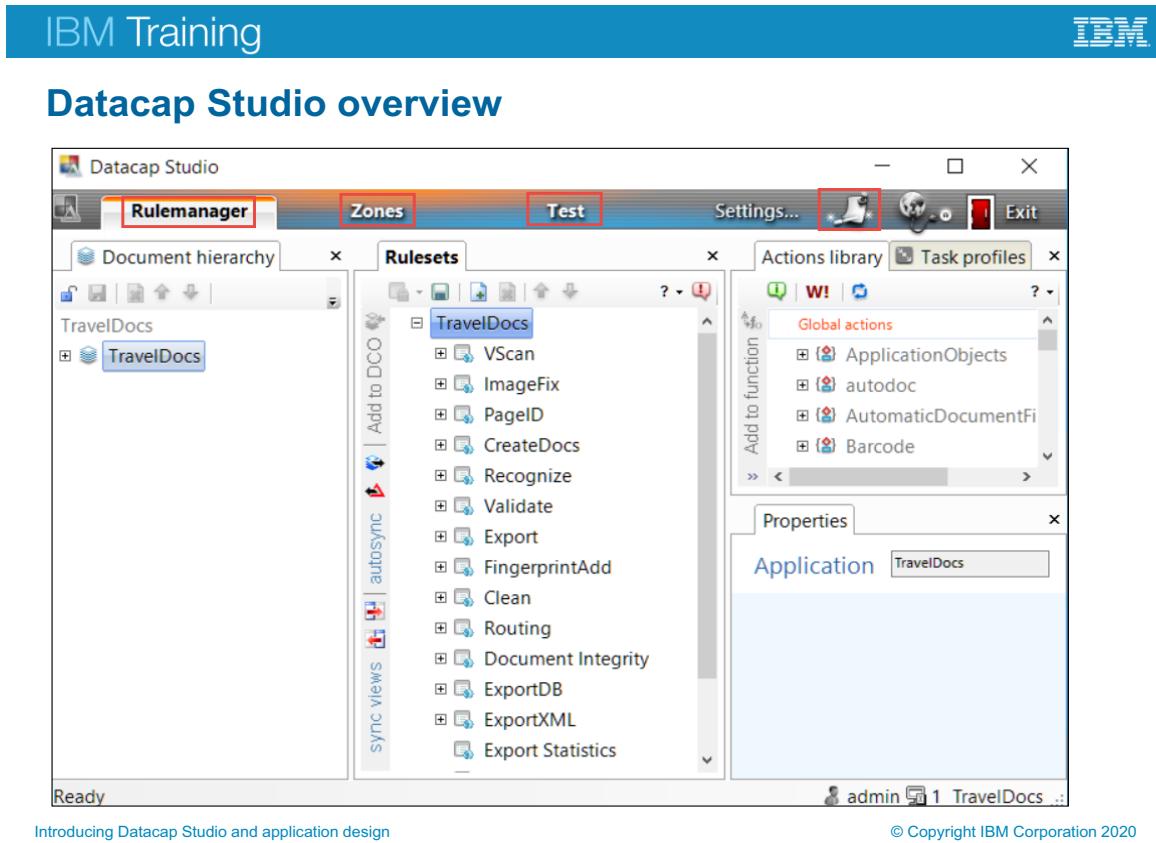


Figure 3-4. Datacap Studio Overview

The screen capture shows the Datacap Studio interface.

Datacap Studio is a rich application development environment for Datacap Capture.

- It includes an Application Wizard that gives you a head start on application development by generating a basic application framework, complete with the supporting folders.
- It provides the tools that you need to develop and test your application.

Datacap Studio contains three main tabs: Rulemanager, Zones, and Test

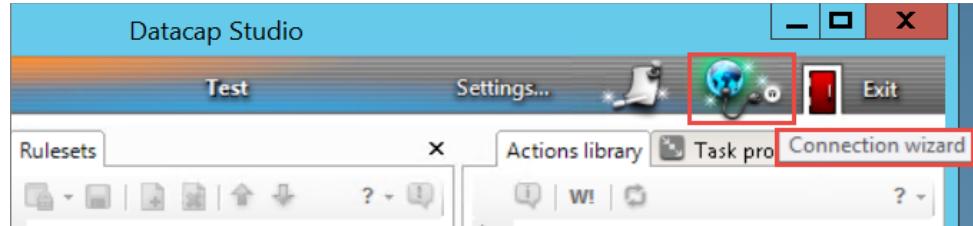
1. In the Rulemanager tab, you can view document structures, rulesets (rules, functions, & actions), action library, properties, and task profiles
2. In the Zones tab, you can add and define fingerprints and view and modify properties of selected objects
3. In the Test tab, you can debug your application, set breakpoints, and step through your application

You learn about the Test tab in a following unit.



Connection wizard in Datacap Studio

- You can switch to a different application within Datacap Studio.



- The wizard opens the Applications dialog box and when you select an application, it prompts you to log in.

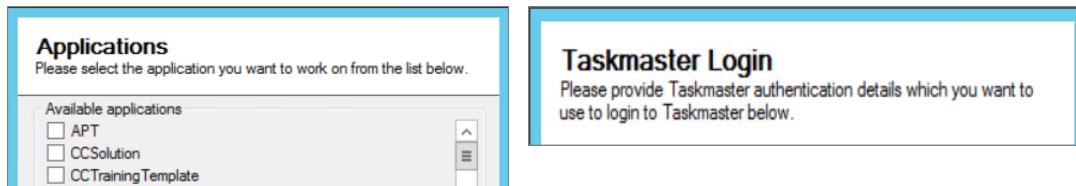
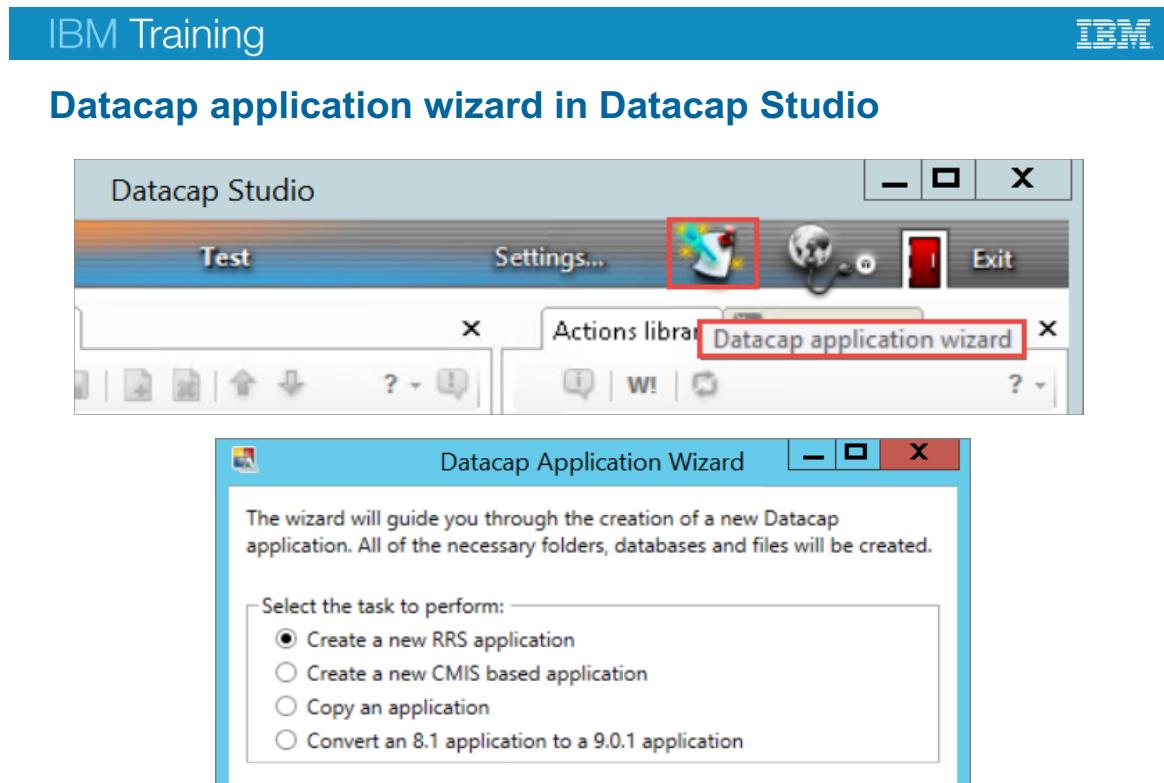


Figure 3-5. Connection wizard in Datacap Studio

The screen capture shows the Connection wizard where you can switch to a different application from within Datacap Studio.

- When you click the Connection wizard icon on the toolbar, the Applications dialog box opens.
- Applications dialog box lists the applications that are in the Datacap.xml file.
- You select an application that you want to open and log in.



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-6. Datacap application wizard in Datacap Studio

The screen capture shows the Datacap application wizard.

- When you click the Datacap application wizard icon on the toolbar, the wizard opens.
- It guides you through the creation of a new Datacap application or copying of an existing application into a new one.

3.2. Datacap Studio - Rulemanager

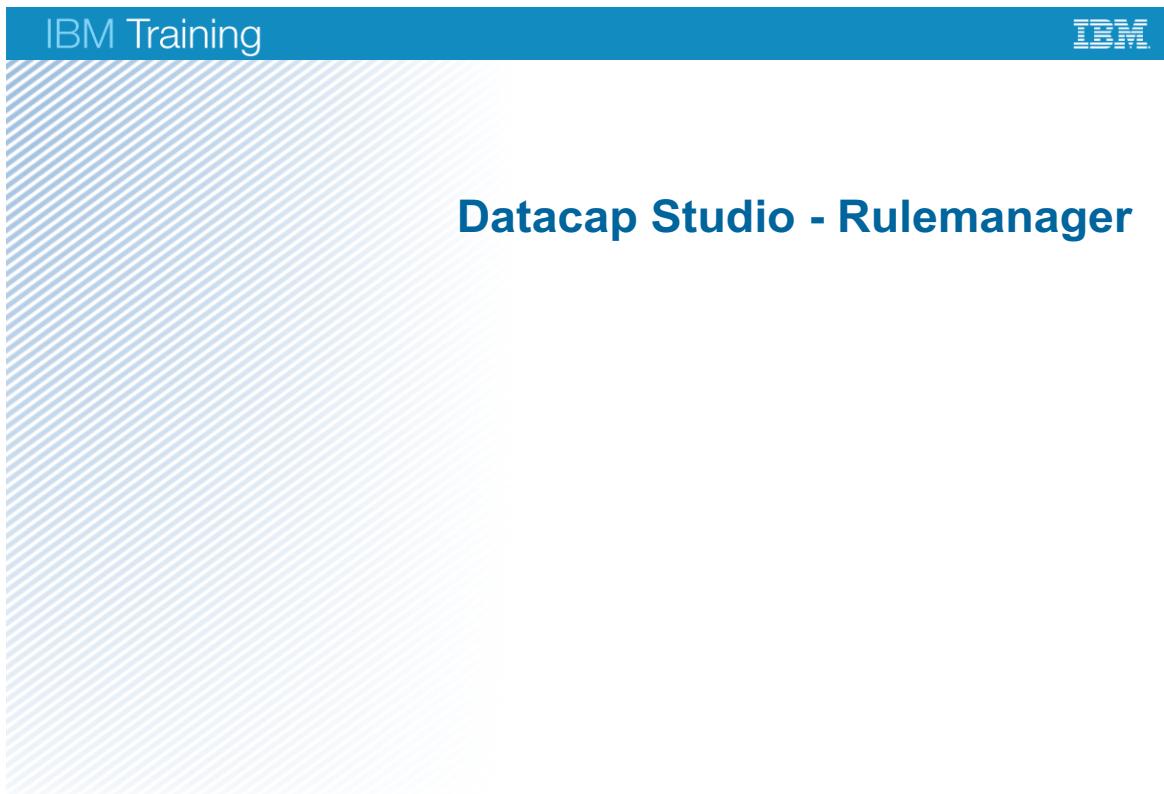
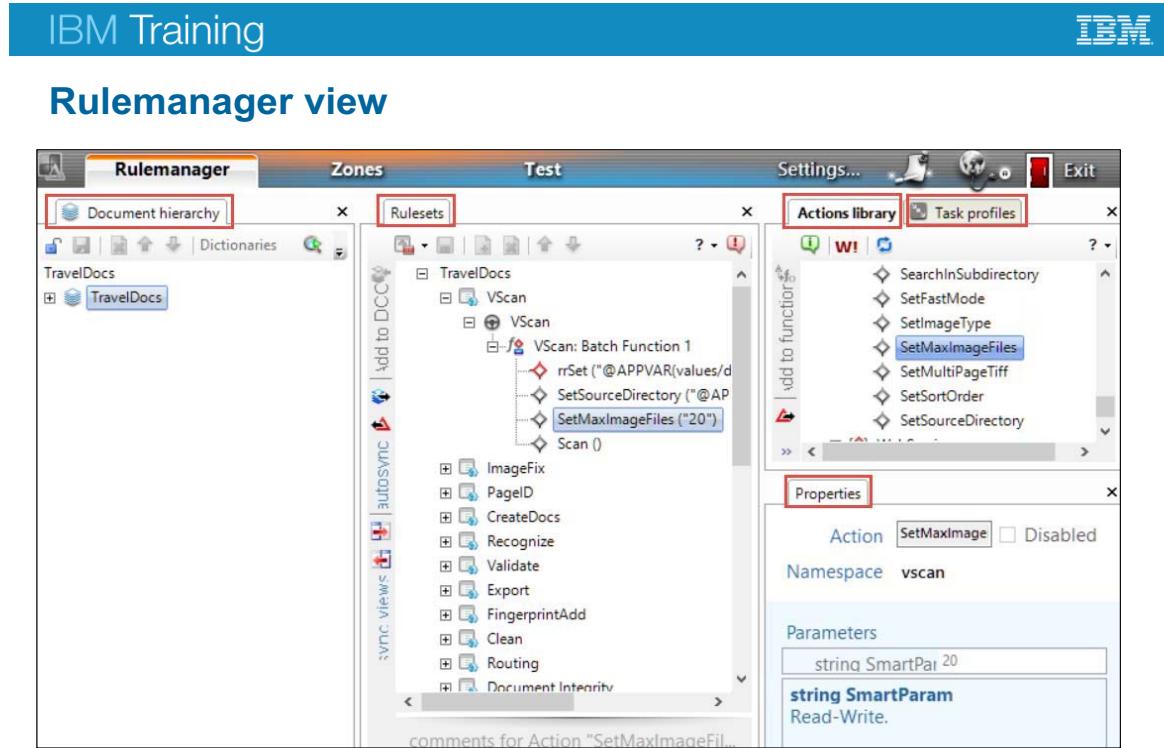


Figure 3-7. Datacap Studio - Rulemanager



[Introducing Datacap Studio and application design](#)

© Copyright IBM Corporation 2020

Figure 3-8. Rulemanager view

The screen capture shows the Rulemanager tab in Datacap Studio that is the primary application development area.

The default layout has the following five panes:

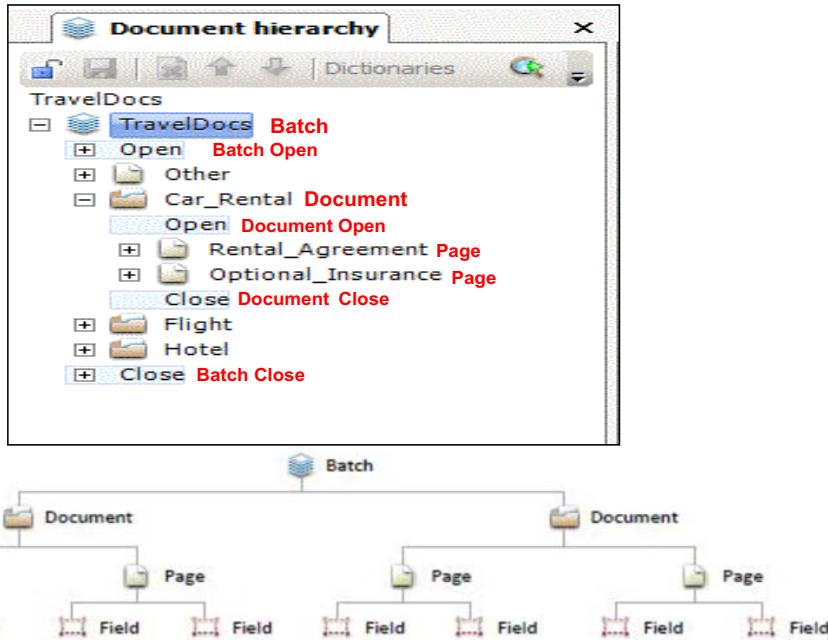
1. Document hierarchy
2. Rulesets
3. Action library
4. Task profiles
5. Properties

You can customize the pane layout by removing or adding panes or by moving them to different locations.

The three main panes in the Rulemanager tab of Datacap Studio represent three aspects of actions:

- Task profiles - When does it happen
- Rulesets - What happens
- Document hierarchy - where does it happen

Document hierarchy (DCO): document and page types



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-9. Document Hierarchy (DCO) - document and page types

The screen capture shows the Document hierarchy pane and a graphical representation of the Batch Structure.

The document hierarchy (also known as the DCO) describes the structure of the documents that your application is designed to handle.

The component levels within the hierarchy are Batch, Document, Page, and Field.

- For each application, batch (Example: TravelDocs) is the root level object.
- Each batch contains one or more document types (Example: Car_Rental, Fight, and Hotel)
- Each document type has at least one page type (Example: Rental_Agreement, Optional_Insurance)

There is always a page type Other for the batch. The page type is used to temporarily assign a page type to all incoming pages until each page is identified in the application.

- Each page type has a number of fields (Example: Pickup_Date)

Open and Close nodes

Each component has an Open and a Close node in the hierarchy. The Open and Close nodes are used to assign processing rules to the beginning or the end of each component.

The next slide contains information on the DCO fields.

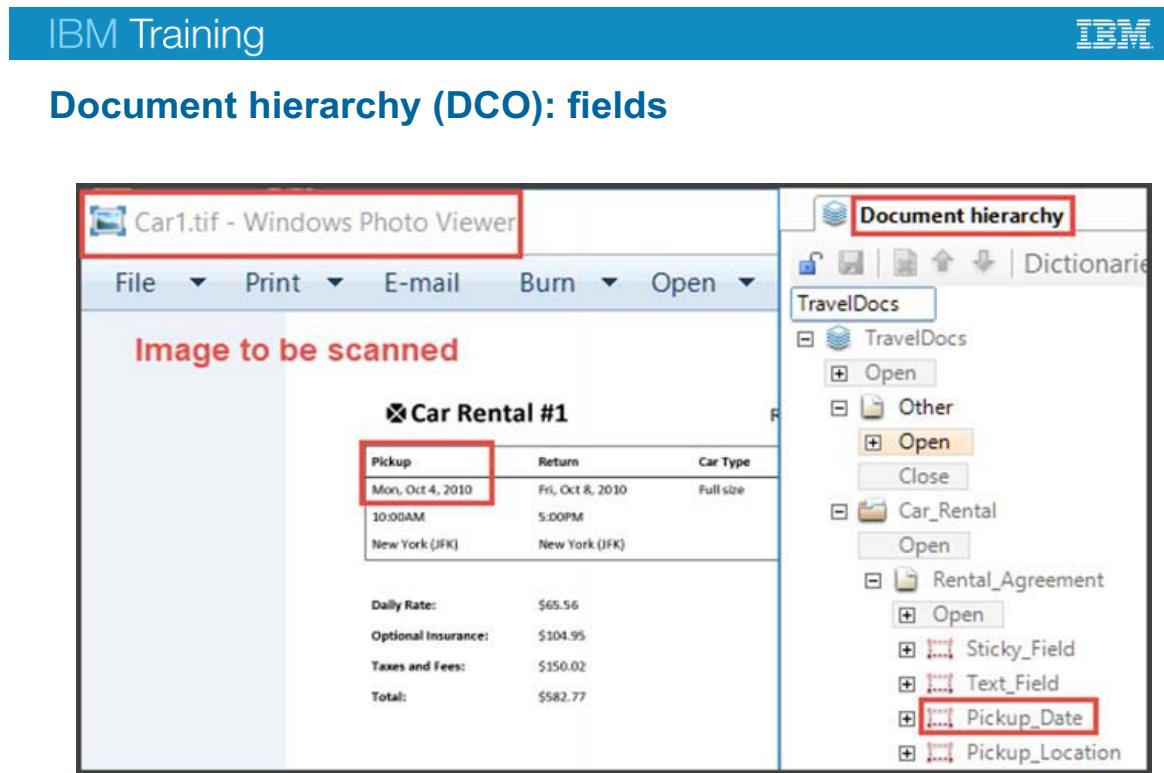


Figure 3-10. Document Hierarchy (DCO) - fields

The screen capture shows the image to be scanned in Windows Photo Viewer on the left and the Document hierarchy pane in Datacap Studio on the right.

Observe that the field on the scan image is mapped to the DCO field (Example: Pickup_Date)

When you create a Datacap application, the wizard creates a folder with that application name. All the configuration information of the application is stored within the folder.

The C:\Datacap\TravelDocs\dco_TravelDocs\TravelDocs.xml file contains the details of Datacap objects (DCO) that you define in Datacap Studio.

Rulesets, rules, functions, and actions

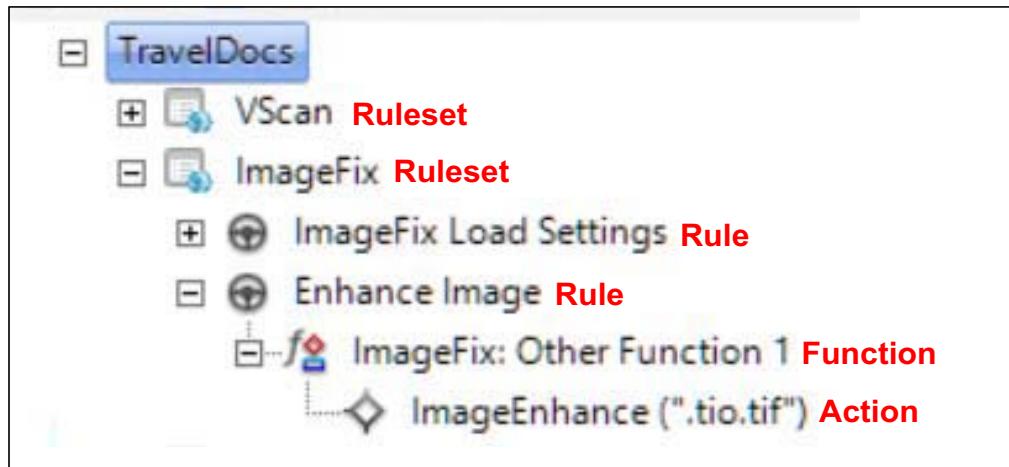


Figure 3-11. Rulesets, rules, functions, and actions

The screen capture shows a sample ruleset and its components: Rules, Functions, and Actions

- Each Datacap application contains many rulesets. For example, ImageFix
- Each ruleset contains one or more rules. For example, Enhance Image
- Each rule contains one or more functions. For example, ImageFix: Other Function 1
- Each function contains one or more actions. For example, ImageEnhance(...)

Rules

Rules are assigned to process specific DCO objects in the document hierarchy (for example, to analyze each page and identify its type). An ordered set of functions (and actions) that are defined in the rule process an object. The rule mapping is discussed in the following slides.

Functions

A function is an ordered list of actions that runs until one of the actions returns a value of False.

Rules are made up of functions that run in the listed order until one of the functions completes or all functions failed.

Actions

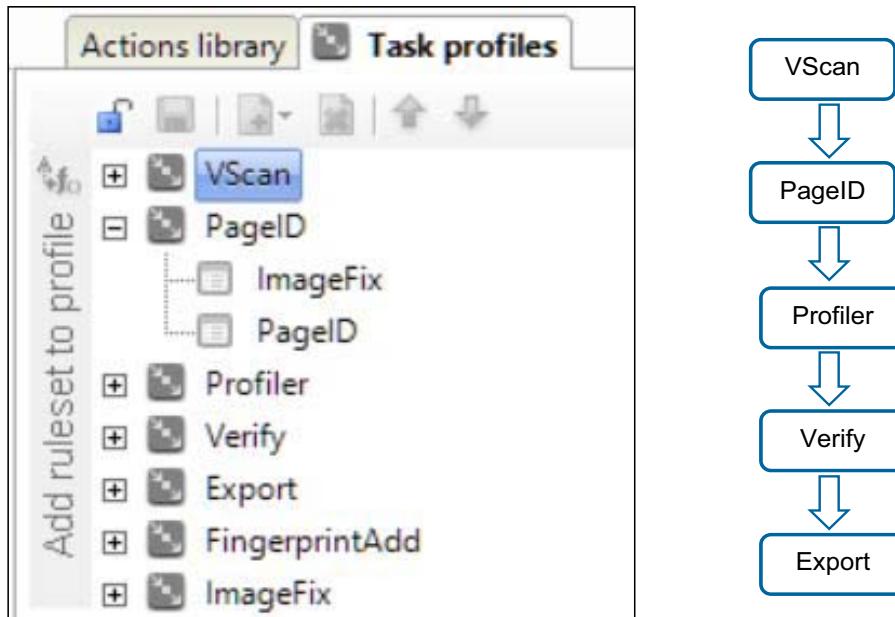
Actions are the fundamental building blocks that are used to define rules. They always return True or False to the Rulerunner service. They are built with .Net or VB script. Actions library provides prebuilt actions. A group of actions forms a function.

When you create a Datacap application, the wizard creates a folder with that application name. All the configuration information of the application is stored within the folder.

Definitions for rulesets are stored in the C:\Datacap\TravelDocs\dco_TravelDocs\rules folder. Each ruleset that is defined in the application has a corresponding <ruleset>.rul file. The .rul files are XML-based.

The collection.xml file contains the list of ruleset names.

Task profiles



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-12. Task Profiles

The screen capture shows the Task profiles pane in the Datacap Studio Rulemanager window.

Task profiles

A workflow consists of a series of tasks and defines a way to process documents.

- Task profiles tab lists the workflow tasks.
- A task profile is made up of a number of rulesets that are arranged in a particular sequence to produce the wanted task processing results.
- They determine the order that rules are processed for a task

A standard workflow for processing documents from Datacap Desktop takes a batch of documents through each of the processing steps.

A job consists of one or more tasks. To process a batch of documents, you must run the batch through each task in the selected job. Some tasks (for example, Export) run without operator intervention, but others (for example, Verify) require an operator.

The job type that you select determines the tasks in the workflow. A typical workflow for a job includes five tasks:

- VScan
- PageID
- Profiler (Extract)
- Verify
- Export

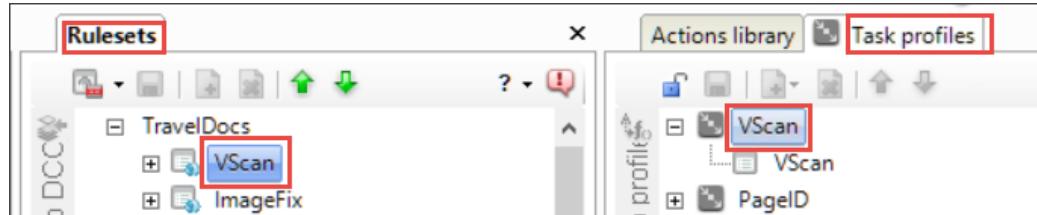
Each task is linked to a task profile and the task profile determines the order that tasks are processed.

Add Rulesets to the Task profile

To add a ruleset to a task in the Task profile, you must select the ruleset in the Rulesets pane. Next, you must select the task in the Task profile pane that you want to be added to the selected task profile. Then, click the 'Add ruleset to profile' link on the left of the Task profiles pane.

IBM Training 

Rulesets and task profiles



[Introducing Datacap Studio and application design](#)

© Copyright IBM Corporation 2020

Figure 3-13. Rulesets and task profiles

The screen capture shows the association of a ruleset to a task profile in the Datacap Studio Rulemanager window.

Add Rulesets to the Task profile

To add a ruleset to a task in the Task profile, you must select the ruleset in the Rulesets pane. Next, you must select the task in the Task profile pane that you want to be added to the selected task profile. Then, click the 'Add ruleset to profile' link on the left of the Task profiles pane.

When you create a Datacap application, the wizard creates a folder with that application name. All the configuration information of the application is stored within the folder.

The C:\Datacap\TravelDocs\dco_TravelDocs\rules\collection.xml file contains the configuration of task profiles.



Figure 3-14. Actions Library

The screen capture shows the Actions library pane of the Datacap Studio Rulemanager window. It also shows the Rulesets tab to demonstrate how actions are used for the rulesets.

Add actions to rules

When building rules in the Ruleset pane, actions from one or more Action libraries can be added to the functions. One or more functions makes a rule.

Actions library

- It provides access to the complete library of prebuilt actions.
- An action is the smallest component of a workflow. It is the equivalent of an instruction in programming terms.
- To get help on an action, select the action and click Information.
- Each item that is listed in the Action library pane is a group of actions of a particular type.
- Click the plus sign next to each group to expand and show all the actions in the group.

Properties tab

- View object properties
- In general the properties pane shows the following items:
 - Object type
 - Object name
 - Object properties, static and configurable (If applicable)

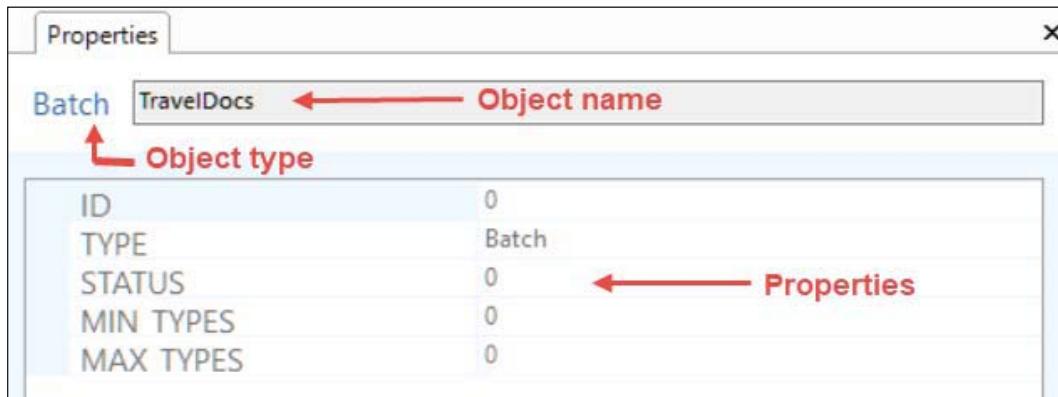


Figure 3-15. Properties tab

The screen capture shows the Properties tab on the Datacap Studio Rulemanager window.

Properties pane

Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters.

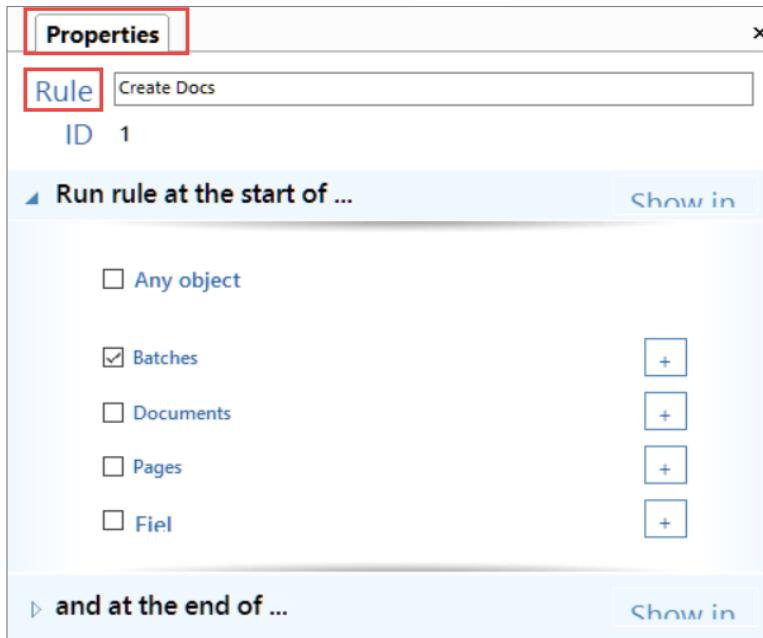
Each object type in the Document hierarchy and the Rulesets pane have a unique representation in the Properties pane:

- Document hierarchy object types are:
 - Application
 - Batch
 - Document
 - Page
 - Field

- Ruleset object type properties are:
 - Application
 - Ruleset
 - Rule
 - Function
 - Action



Rule object properties tab



[Introducing Datacap Studio and application design](#)

© Copyright IBM Corporation 2020

Figure 3-16. Rule object properties tab

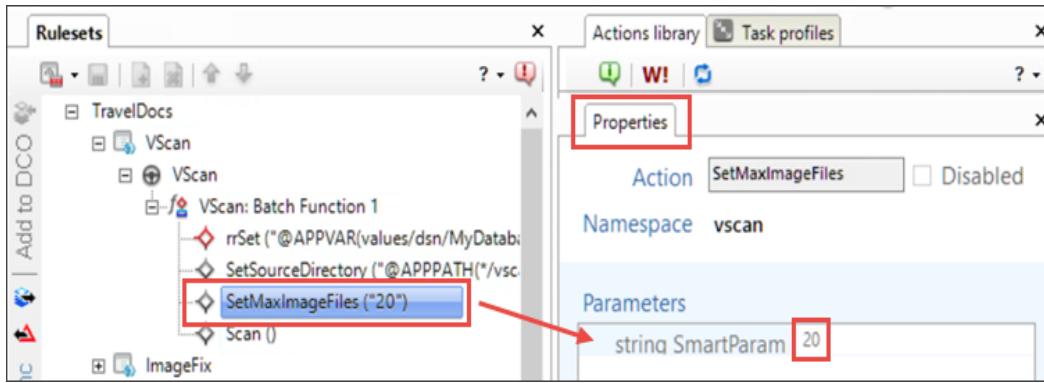
The screen capture shows Properties pane of the Datacap Studio window when a rule is selected in the Ruleset pane. It shows the controls that are used to associate rules with the DCO or Batch Structure objects.

- Mapping at beginning or end of object rule processing
 - Mapping configuration done under 'Run rule at the start of ...' is equivalent to mapping a rule to the Open node in the conventional method.
 - Mapping configuration that is done under 'and at the end of ...' is equivalent to mapping a rule to the Close node in the conventional method.
- Mapping to all objects or all objects of a specific type
 - The 'Any object' checkbox causes the selected rule to associate with all DCO objects.
 - The 'Batch' checkbox causes the selected rule to associate with all batch level objects.
 - The 'Documents' checkbox causes the selected rule to associate with all document level objects.
 - The 'Page' checkbox causes the selected rule to associate with all page level objects.
 - The 'Field' checkbox causes the selected rule to associate with all Field level objects.
- Mapping to a specific object of a specific type

IBM Training



Properties tab



[Introducing Datacap Studio and application design](#)

© Copyright IBM Corporation 2020

Figure 3-17. Properties tab

The screen capture shows the Properties tab on the Datacap Studio Rulemanager window.

View object properties

In general the properties pane shows the following items:

- Object type
- Object name
- Object properties, static and configurable (If applicable)

Properties pane

Shows the properties for the selected document hierarchy or ruleset object. If the corresponding pane is locked for editing, you can also modify existing properties, including specifying action parameters.

Each object type in the Document hierarchy and the Rulesets pane have a unique representation in the Properties pane:

- Document hierarchy object types are:
 - Application
 - Batch
 - Document
 - Page
 - Field
- Ruleset object type properties are:
 - Application
 - Ruleset
 - Rule
 - Function
 - Action



Action object properties tab

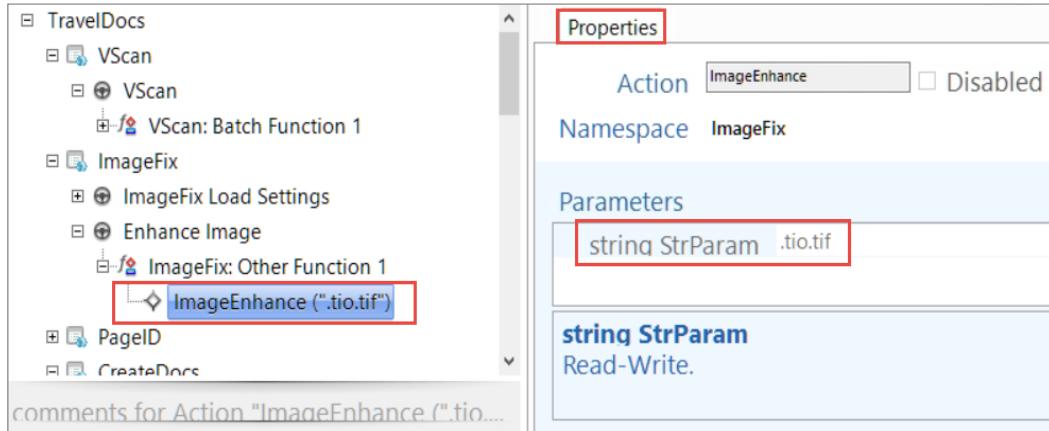


Figure 3-18. Action object properties tab

The screen capture shows Properties pane of the Datacap Studio window when an action is selected in the Ruleset pane. It shows the parameters for an action.

The properties pane is where you go to set the parameters that are passed to actions.

3.3. Datacap Studio - Zones

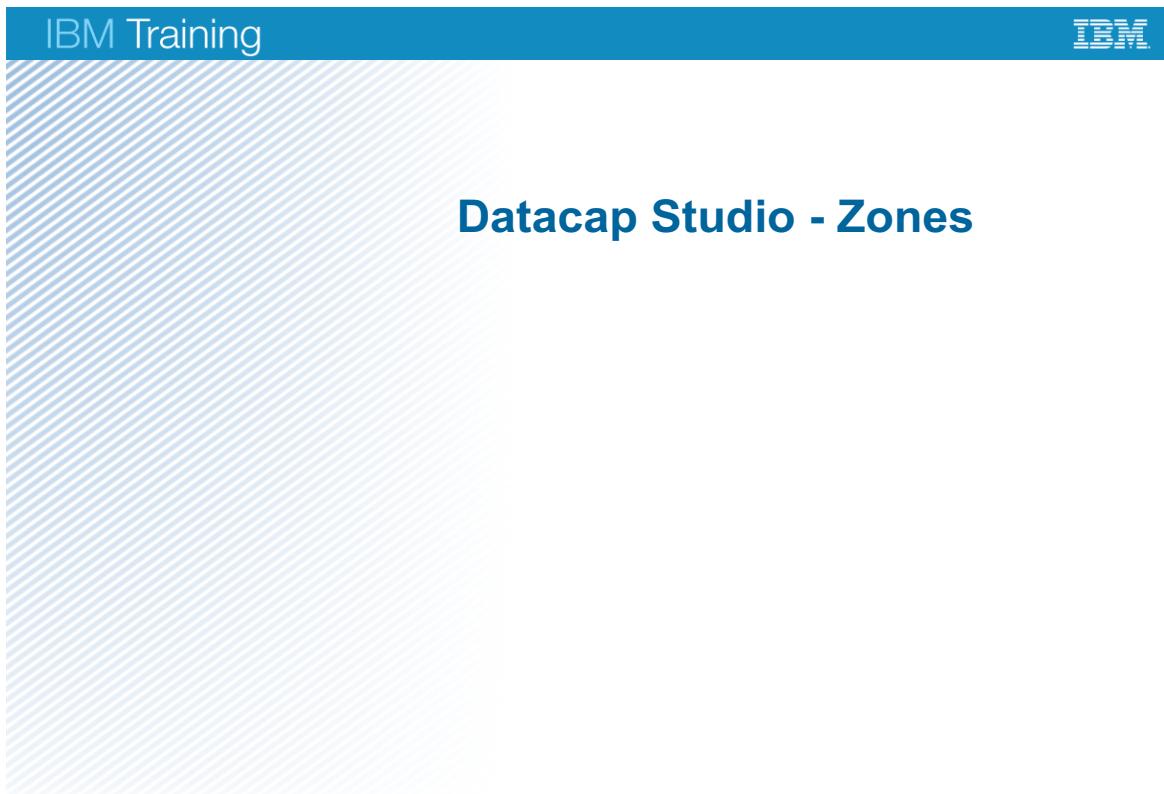
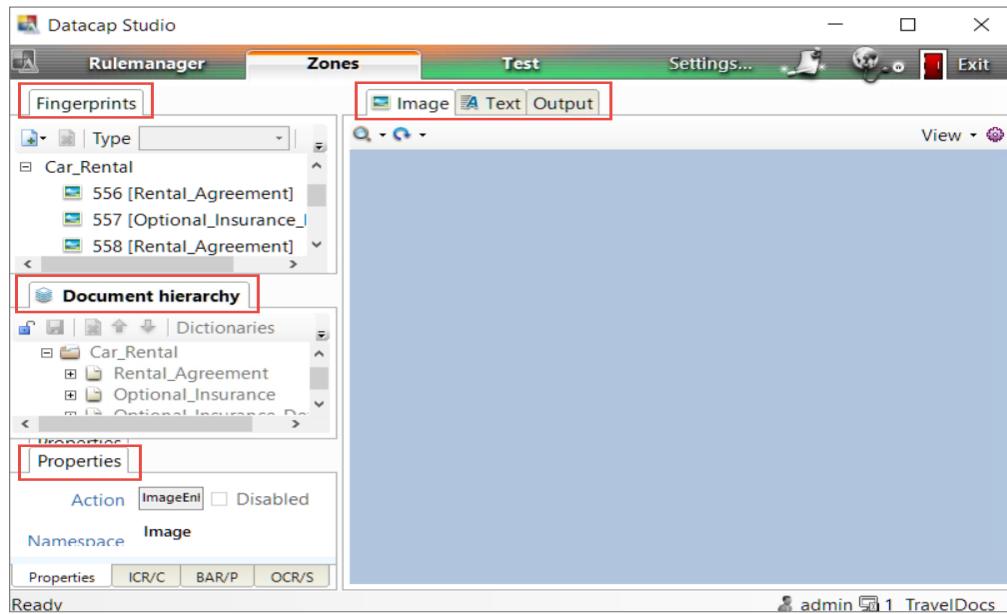


Figure 3-19. Datacap Studio - Zones

IBM Training



Zones view



[Introducing Datacap Studio and application design](#)

© Copyright IBM Corporation 2020

Figure 3-20. Zones view

The screen capture shows the Zones tab in Datacap Studio. You define fingerprints and view and modify properties of selected objects.

The default layout has the following four panes:

1. Fingerprints
2. Document hierarchy
3. Properties
4. Image

Fingerprints

The pane shows the application fingerprint library and you can add fingerprints for new page types.

Document hierarchy

Defines the structure of the documents that you are processing and how each element within the structure is processed.

Properties

You can manage the properties for the selected document hierarchy object.

In the Properties pane, you specify recognition options for the selected object.

Datacap Capture supports multiple recognition engines. Panes for ICR/C, BAR/P, and OCR/S are shown by default.

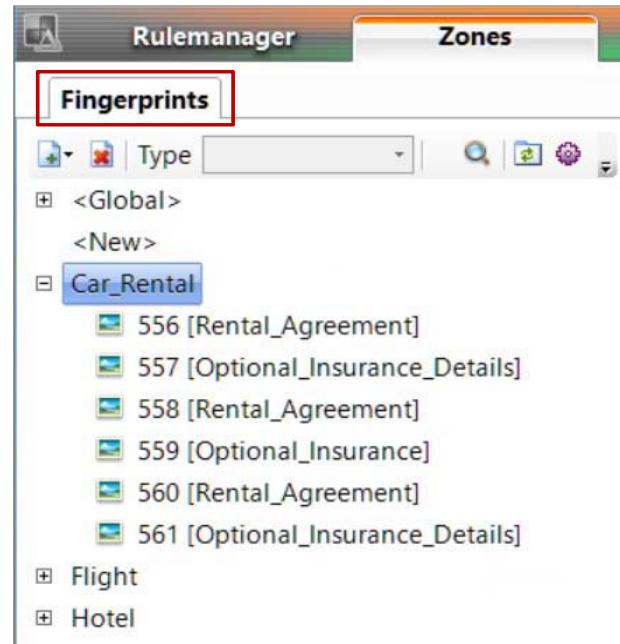
You can access other panes by right-clicking within the Properties panel and selecting Show panes.

Image view

Shows the selected fingerprint image and any recognition zones. This view is where you draw new recognition zones. If you created the fingerprints with full page recognition, you can view the recognition results in the Text pane.

Fingerprints pane

- Fingerprint management
- Displays the application's fingerprints library
- Fingerprints classes and fingerprints can be added and deleted from this view



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-21. Fingerprints pane

The screen capture shows the Fingerprints pane in Zones tab.

Document hierarchy pane

- Describes the structure of the documents
- Displays how each element in the structure is related to the other elements

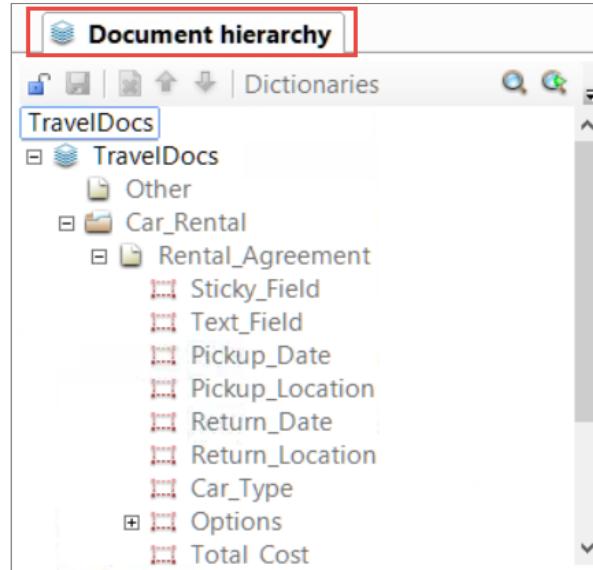


Figure 3-22. Document hierarchy pane

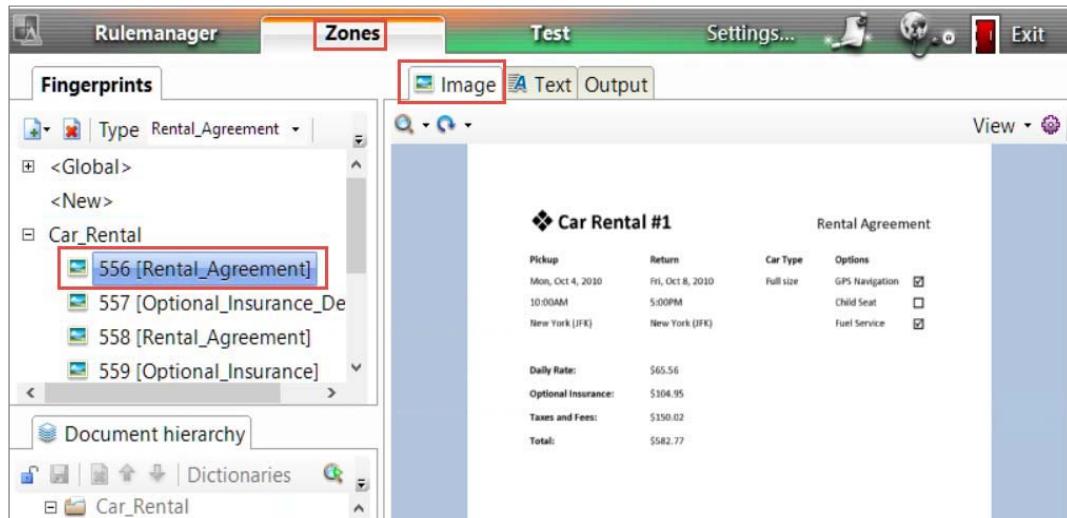
The screen capture shows the Document hierarchy pane in Zones tab.

The document hierarchy describes the structure of the documents that your application is designed to process. The levels within the hierarchy are batch, document, page, and field.

It also displays how each element in the structure is related to the other elements and how each object is processed.

IBM Training

Image view



Introducing Datacap Studio and application design

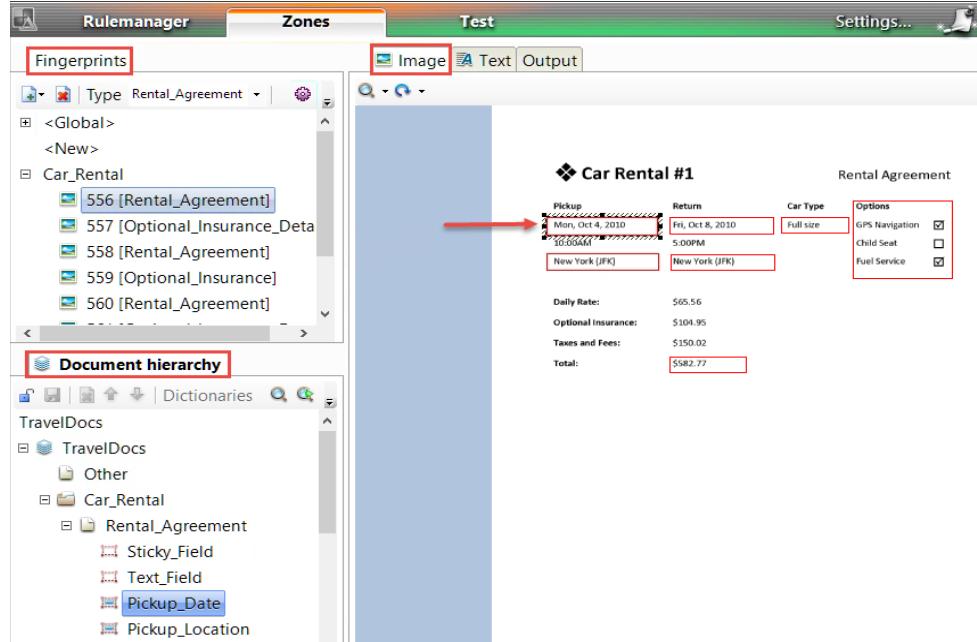
© Copyright IBM Corporation 2020

Figure 3-23. Image View

The screen capture shows the Image pane in Zones tab.

IBM Training

Zones



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-24. Zones

The screen capture shows the zones for the fields on the Image pane in Zones tab.

The Image tab displays the selected fingerprint image and any recognition zones. You can draw new zones or modify the existing zones.

If you create fingerprints by using full page recognition, you can view the recognition results in the Text tab.

3.4. Datacap application design



Figure 3-25. Datacap application design

Steps to design and configure an application

- Do an end-to-end evaluation of the batch process to be automated
- Analyze the data source
- Implement a sandbox system
- Build the Datacap application
- Test and adjust the application

Figure 3-26. Steps to design and configure an application

Do a complete end-to-end evaluation

- Analyze the end-to-end business process that is associated with your corporate documents that need to be captured.
- Determine of what data is going to be captured so you know what to look for in the documents.
 - How are the images to be captured including resolution and color properties of the images.
 - How to locate and extract the data?
 - How to validate the data?
 - Where is the output to be sent?
 - Does it need any special format?
- Determine the steps people must do.
- At what point can the documents be removed from the system or archived?

Figure 3-27. Do a complete end-to-end evaluation

Scenario for a sample application

- An insurance company needs to extract data from a Claims document that is on the file system folder.
- For the following fields:
 - Driver name
 - Policy Number
 - Vehicle License
- Document
 - Classify type
- Fields
 - Extraction type

Auto Insurance Claim Form
Insurance Company A

Policy Holder Address :	Chris Joes, 13 Johan Street NY 987654
Driver Name :	Chris Joes
Policy Number :	48998813
Incident Number :	CL-4328941
Incident Date:	03/03/11
Incident Time:	14:15
Vehicle License :	2296RG
Vehicle Colour :	Black
Vehicle Manufacturer :	Tos
Vehicle Model :	Quaser
Year of Manufacture :	2009
Chassis Number (VIN) :	2C5BB4CRX82LP3489

Incident Description:
Incident occurred on Highway 69, NY.
Damage occurred to rear of the vehicle. Damaged right tail light and bumper.

Was anyone injured in this incident which required medical attention? Yes No

By signing the adjacent box and entering the current date, you agree the information above is accurate to the best of your knowledge.

Signature: *Chris Jones*
Dated: 03/03/2011

Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-28. Scenario for a sample application

An insurance company needs to extract data from a Claims document that is on the file system folder and you want to automate the process with Datacap.

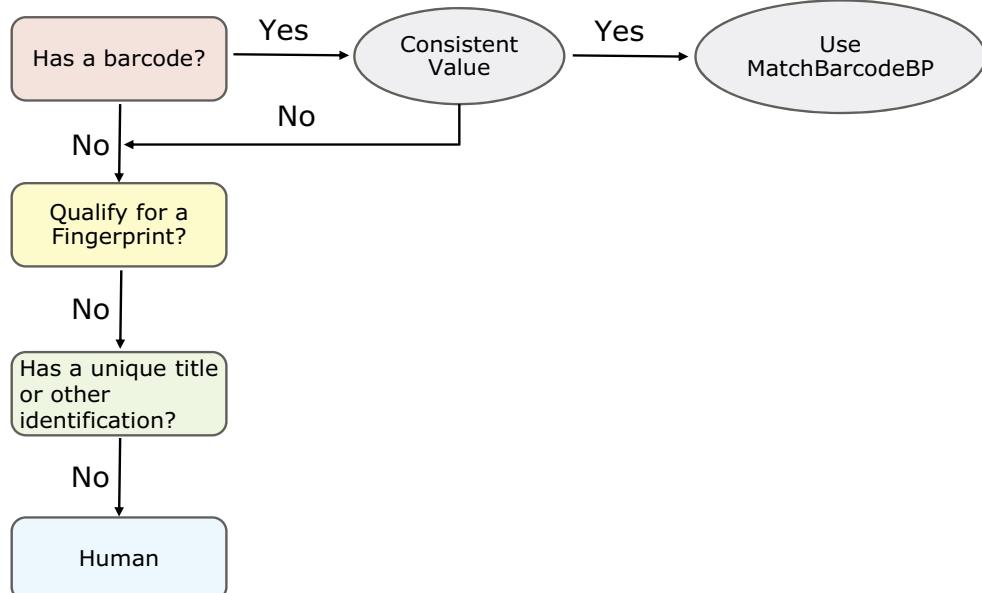
In this example, you extract values for the following fields:

- Driver name
- Policy Number
- Vehicle License

In the following slides, you decide what classify type to use for the document and what extraction type to use for the fields.

Deciding which method to use for document classification

- Guides you through the creation of a new Datacap application.



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-29. Deciding which method to use for document classification

For deciding which method to use for document classification, you ask the following questions:

- a. Does the page have a barcode?
- b. Does it qualify for a Fingerprint?
- c. Does the page have a unique title or other identification?

If you have more than one barcode on a single page, you get all the barcode values. Then, check the value and select the one that identifies the page (has a constant value).

Planning the workflow details for the application

- Page ID
 - Classify the document type
 - Use Barcode on the page
- Profiler - Extract
 - OCR - simple line based
 - Fields can be extracted with fingerprint zones
 - Standard workflow
- Validate
 - One user review to validate the data
- Export
 - Simple file export to a text file
 - To an export folder on the file system

Auto Insurance Claim Form
Insurance Company A

Policy Holder Address :	Chris Joes, 13 Johan Street NY 987654
Driver Name :	Chris Joes
Policy Number :	48998813
Incident Number :	CL-4328941
Incident Date:	03/03/11
Incident Time:	14:15
Vehicle License :	2296RG
Vehicle Colour :	Black
Vehicle Manufacturer :	Tos
Vehicle Model :	Quaser
Year of Manufacture :	2009
Chassis Number (VIN) :	2C5BB4CRX82LP3489

Incident Description:
Incident occurred on Highway 69, NY.
Damage occurred to rear of the vehicle. Damaged right tail light and bumper.

Was anyone injured in this incident which required medical attention? Yes No

By signing the adjacent box and entering the current date, you agree the information above is accurate to the best of your knowledge.

Signature: *Chris Jones*
M M D D Y Y
Dated: 03 03 2011
CLAIM-29A

Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-30. Planning the workflow details for the application

For your application, you review the document and plan the Datacap workflow tasks.

In this example, assume the following details:

- The field positions on the page are constant for all the forms of this type.
- The value for the upper bar code varies based on the values on the form.
- The value for the lower bar code remains constant and it is unique for the form.

For this document, you can use the lower bar code to identify the page. It is unique to the page and has a constant value (*CLAIM-29A*).

If a page has a bar code and it has a constant value, always select bar code method. Because the confidence level for bar code is high.

Since the fields positions are constant for this image, you can use fingerprints and zones to extract the value from the fields.

Planning the task details for the workflow

- VScan
 - VScan
- PageID
 - Barcode
 - OCR-S
 - Fingerprint
- Profiler
 - ReadZones FPXML
 - Validate
- Verify
 - Validate
- Export
 - Export all fields to text file

Auto Insurance Claim Form
Insurance Company A

Policy Holder Address :	Chris Joes, 13 Johan Street NY 987654
Driver Name :	Chris Joes
Policy Number :	48998813
Incident Number :	CL-4328941
Incident Date:	03/03/11
Incident Time:	14:15
Vehicle License :	2296RG
Vehicle Colour :	Black
Vehicle Manufacturer :	Toa
Vehicle Model :	Quaser
Year of Manufacture :	2009
Chassis Number (VIN) :	2C5BB4CRX82LP3489
Incident Description: Incident occurred on Highway 69, NY. Damage occurred to rear of the vehicle. Damaged right tail light and bumper.	
Was anyone injured in this incident which required medical attention? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No	
By signing the adjacent box and entering the current date, you agree the information above is accurate to the best of your knowledge. Signature:  Dated: 	

Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-31. Planning the task details for the workflow

The actions for the VScan depend on the source for the document that is scanned.

Examples: Scanner, email, mobile

The fields for the page are set up exactly as shown in the screen capture and does not change.

OCR-S is added under the PageID task (rather than Extract) so that the fingerprint has the CCO.

Steps to build an application

Steps	Datacap tools
• Copy or create an application	Datacap Studio
• Configure jobs and workflow	Datacap Navigator
• Complete application settings	Application Manager
• Create Rulerunner threads	Rulerunner Manager

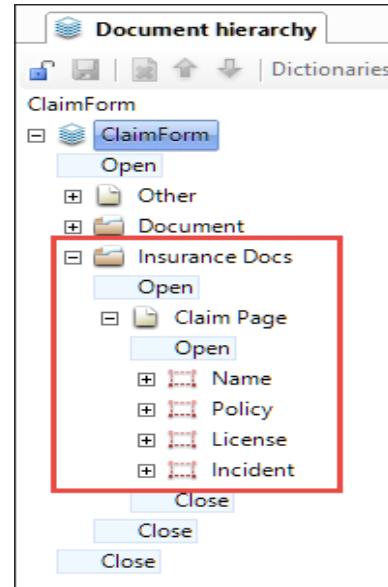
Figure 3-32. Steps to build an application

Steps to build an application

1. In Datacap Studio, create a copy of an existing application or create an application based on the templates available (Forms or Learning templates)
2. In Datacap Navigator admin tool, configure workflow and jobs.
3. In Datacap Application Manager, complete the application settings
4. In Datacap Rulerunner Manager, create rulerunner threads to run the background tasks.

Set up Document hierarchy in Datacap Studio

- Add Document object
 - Example: Insurance Docs
- Add Page object
 - Example: Claim Page
- Add Field objects
- Examples:
 - Name
 - Policy
 - License
 - Incident



[Introducing Datacap Studio and application design](#)

© Copyright IBM Corporation 2020

Figure 3-33. Set up Document hierarchy in Datacap Studio

In Datacap Studio > Rulemanager > Document hierarchy tab, you create Datacap objects (DCO) for your application.

- Each application has a root level Batch object. To which you add one or more Document objects.

Example: Insurance Docs

- To each Document object, you add one or more Page objects

Example: Claim Form

- To each Page object, you add one or more Field objects

Examples: Name, Policy, License, and incident

Order of operations in Datacap

- (B) Open is run first
- (D) Open is run next
 - (P) Open is run third
 - (F) Open is run fourth
 - (F) Close
 - (P) Close
 - (D) Close
- (B) Close

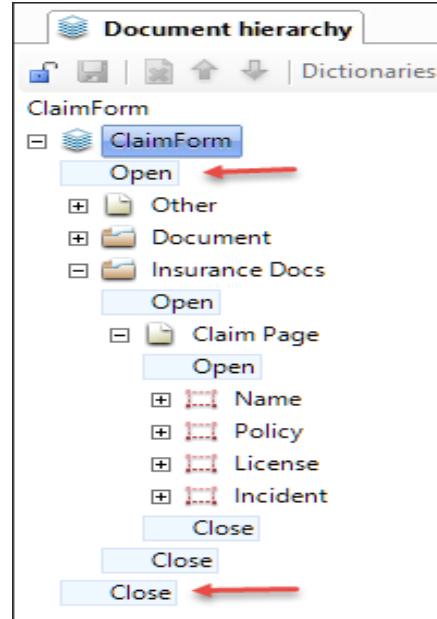


Figure 3-34. Order of Operations in Datacap

The actions that you define in the Rulesets are associated with the Document hierarchy (DCO).

The slide shows the order of operation for a batch of a Datacap application.

- (B) - Batch
- (D) - Document
- (P) - Page
- (F) - Field

Rulesets and task profiles for the application

Rulesets	Rulesets (Continued)	Task Profiles
<ul style="list-style-type: none"> • VScan <ul style="list-style-type: none"> ▪ (B) Open <ul style="list-style-type: none"> - Import • PagelD <ul style="list-style-type: none"> ▪ (P) Open <ul style="list-style-type: none"> - Barcode - OCR - Fingerprint • Create Docs <ul style="list-style-type: none"> ▪ (B) Open <ul style="list-style-type: none"> - Create Docs ▪ (P) Open <ul style="list-style-type: none"> - Create Fields 	<ul style="list-style-type: none"> • Recognize <ul style="list-style-type: none"> ▪ (P) Open <ul style="list-style-type: none"> - ReadZonesFPXML • Validate <ul style="list-style-type: none"> ▪ (F) Open <ul style="list-style-type: none"> - Validate rules • Export <ul style="list-style-type: none"> ▪ (B) Open <ul style="list-style-type: none"> - Set export parameters - ExportAllFields() 	<ul style="list-style-type: none"> • VScan <ul style="list-style-type: none"> ▪ VScan • PagelD <ul style="list-style-type: none"> ▪ PagelD ▪ Create Docs • Profiler <ul style="list-style-type: none"> ▪ Recognize ▪ Validate • Verify <ul style="list-style-type: none"> ▪ Validate • Export <ul style="list-style-type: none"> ▪ Export • FingerprintAdd

Figure 3-35. Rulesets and Task profiles for the application

The columns on the slide show Rulesets for the sample application and Task profiles with the rulesets that are associated with them.

Rulesets:

- Rulesets are associated with Datacap objects.
- Open means to do the action at the beginning and Close means do the action at the end.
- When you start a ruleset, you can "Open" whatever level of the object that you need. It does not have to be Batch always. For example, you can directly open a page.
- But you cannot have two operations at the same level. For example, you cannot have two (P) Open

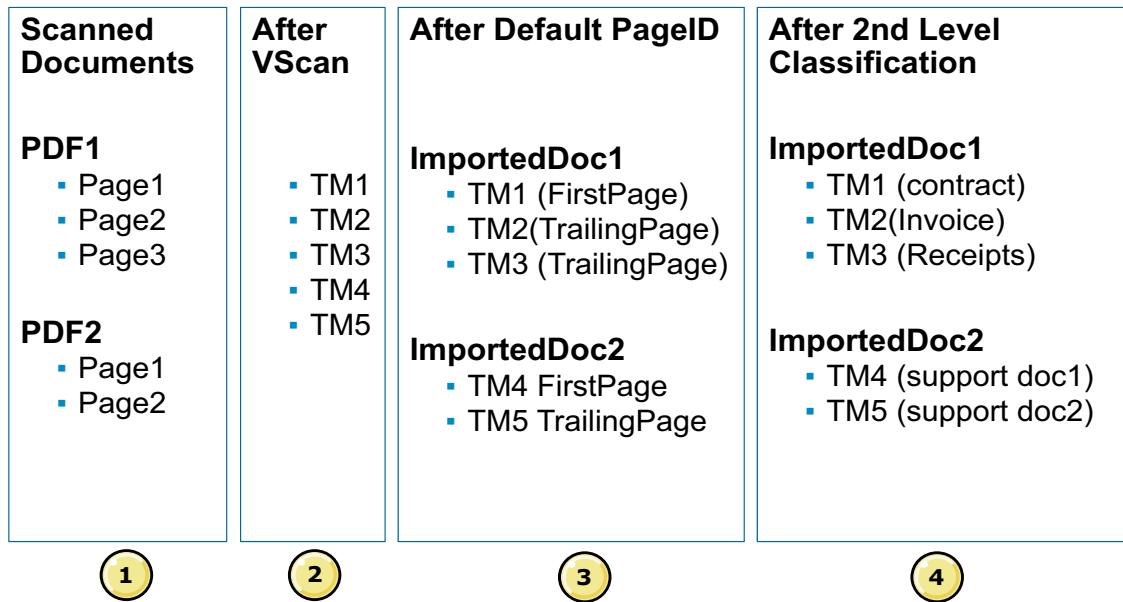
Why the ReadZonesFPXML is at the page level?

ReadZonesFPXML extracts data from the fields. However, it is added to the page level.

Task Profiles

FingerprintAdd must be added manually to the Task Profiles.

Example of Page ID and document classification



Introducing Datacap Studio and application design

© Copyright IBM Corporation 2020

Figure 3-36. Example of Page ID and document classification

The slide shows an example of Page ID and document classification process.

1. A set of sample documents that are used for scanning
2. The same documents after they are scanned, they are listed as individual pages
3. Pages are identified and documents are created
4. Pages in the document are classified

Unit summary

- Identify the application development features in Datacap Studio
- Describe the Datacap application design concepts

Figure 3-37. Unit summary

Review questions

1. True or False: Datacap Studio provides access to the Application Wizard through an icon on the toolbar.
2. True or False: The order of Rulesets in the ruleset pane determines the order in which rules are run.
3. True or False: Datacap Studio is the capture job processing environment.



Figure 3-38. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. True or False: Datacap Studio provides access to the Application Wizard through an icon on the toolbar.
The answer is True.

2. True or False: The order of rulesets in the ruleset pane determines the order in which rules are run.
The answer is False.

3. True or False: Datacap Studio is the capture job processing environment.
The answer is False.



Figure 3-39. Review answers

2. The order of rulesets in the task profiles determines the order in which rules are executed.
3. Datacap Studio is the application development environment.

Unit 4. Building a Datacap application with Forms Template

Estimated time

00:45

Overview

In this unit, you learn how to build a Datacap application with Forms Template in Datacap Studio.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacapoc.doc/datacap_9.1.7.htm

Unit objectives

- Describe how to create a Datacap application with Forms Template
- Learn the application configuration in IBM Content Navigator
- Identify the application settings in Datacap Application Manager
- Describe document hierarchy
- Learn about the fingerprints and zones configuration

Figure 4-1. Unit objectives

Topics

- Create a Datacap application with Forms Template
- Configure the application in IBM Content Navigator
- Edit the application settings in Datacap Application Manager
- Define document hierarchy
- Configure fingerprints and zones

Figure 4-2. Topics

4.1. Create a Datacap application with Forms Template

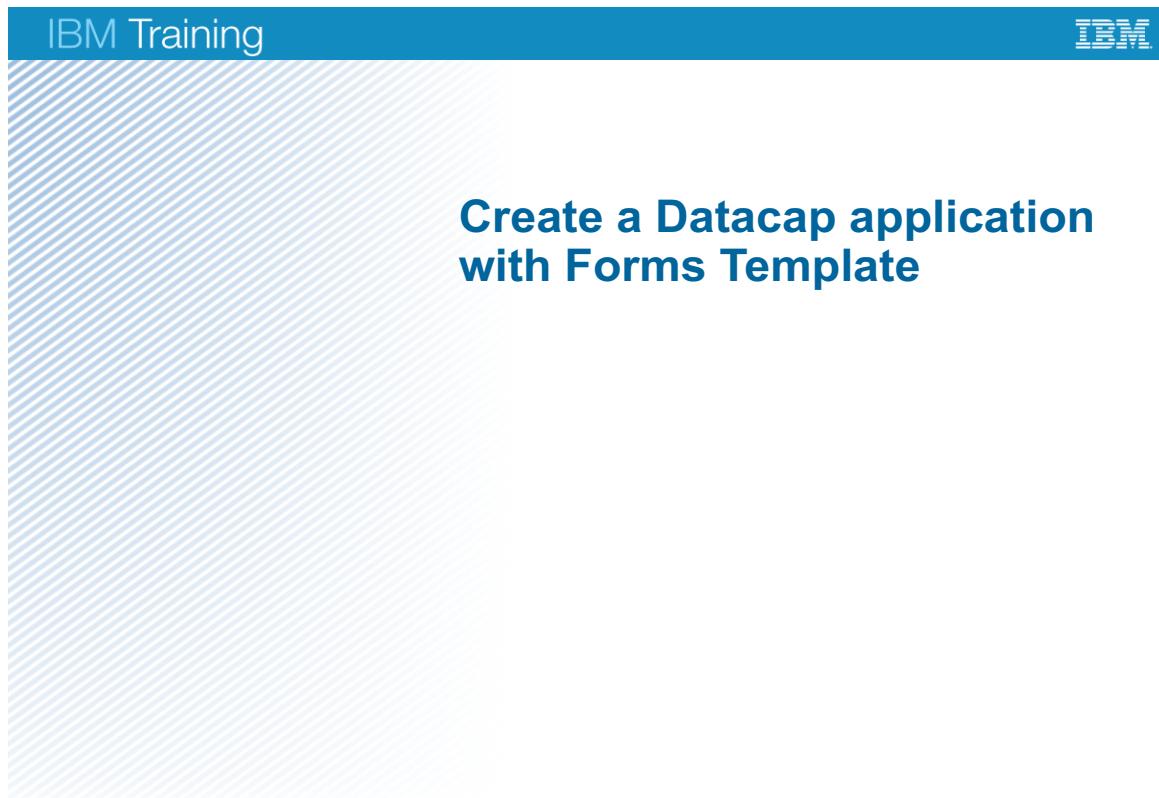


Figure 4-3. Create a Datacap application with Forms Template



Scenario for the application

- An insurance company needs to extract data from a Claims document that is on the file system folder.
- For the following fields:
 - Driver name
 - Policy Number
 - Vehicle License
 - Incident location
- Document
 - Classify type - Barcode
- Fields
 - Extraction type - Fingerprint & zoning
 - Locate rules

Auto Insurance Claim Form
Insurance Company A

Policy Holder Address :	Chris Joes, 13 Johan Street NY 987654
Driver Name :	Chris Joes
Policy Number :	48998813
Incident Number :	CL-4328941
Incident Date:	03/03/11
Incident Time:	14:15
Vehicle License :	2296RG
Vehicle Colour :	Black
Vehicle Manufacturer :	Tos
Vehicle Model :	Quaser
Year of Manufacture :	2009
Chassis Number (VIN) :	2C5BB4CRX82LP3489

Incident Description:
Incident occurred on Highway 69, NY.
Damage occurred to rear of the vehicle. Damaged right tail light and bumper.

Was anyone injured in this incident which required medical attention? Yes No

By signing the adjacent box and entering the current date, you agree the information above is accurate to the best of your knowledge.

Signature: *Chris Jones*
Dated: *03/03/2011*

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-4. Scenario for the application

In the previous unit, you learned about the application design. In this unit, you learn how to use the application design and build an application.

An insurance company needs to extract data from a Claims document that is on the file system folder and you want to automate the process with Datacap.

Document type: Insurance Docs

Page type: Claim Page

In this example, you extract values for the following fields:

- Driver name
- Policy Number
- Vehicle License

Steps to build an Datacap application

Steps	Datacap tools
• Copy or create an application	Datacap Studio
• Configure jobs and workflow	Datacap Navigator
• Complete application settings	Application Manager
• Define DCO, create rulesets, configure task profiles, fingerprints, and zones for the application	Datacap Studio
• Create Rulerunner threads	Rulerunner Manager

Figure 4-5. Steps to build an Datacap application

Steps to build an application

1. In Datacap Studio, create a copy of an existing application or create an application based on the templates available (Forms or Learning templates)
2. In Datacap Navigator admin tool, configure workflow and jobs.
3. In Datacap Application Manager, complete the application settings
4. In Datacap Studio, configure the application (Define document hierarchy, create rulesets, configure task profiles, configure fingerprinting & zones, and test the application)
5. In Datacap Rulerunner Manager, create rulerunner threads to run the background tasks.

Datacap application templates

- A template is an advanced framework for creating an application
 - They are functional applications
- Datacap provides two templates:
 - Forms template for images with structured data
 - Learning template for dynamic forms like invoices
- Developers can create their own templates

Figure 4-6. Datacap application templates

Templates are functional applications with:

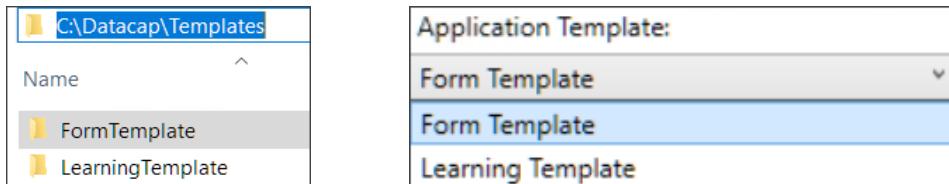
- A built-in basic document hierarchy structure
- A batch workflow that is made up of the basic task profiles
- Commonly used configurable rulesets that are already built into the task profiles

Save your own custom template.

- If you have a base application that you routinely use for creating new applications, you can save it as a template them in the C:\Datacap\templates folder.
- By placing an application in the templates folder, you enable it to be selected from a template list when you create an application with Datacap Application Wizard.
- Your templates are just another application but it is good practice to never access them like an application with any of the development and production tools.
- Always use the Application Wizard to copy and rename, then access the copy.

Template folder structure

- A template has the same folder structure as any other applications.
 - batches
 - dco_<Template Name>
 - export
 - fingerprints
 - images
- Template definitions are in the C:\Datacap\Templates folder.
 - The names of the templates from this folder populate the list to choose from in the Application Wizard



Building a Datacap application with Forms Template

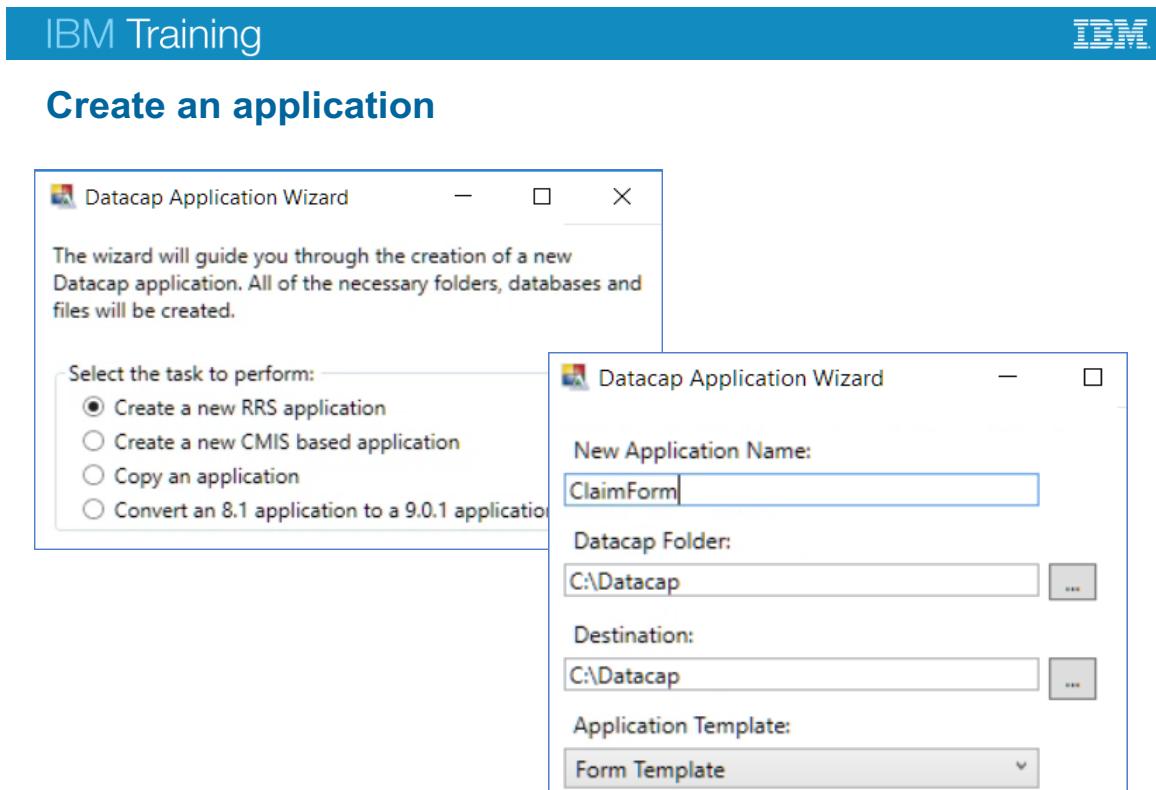
© Copyright IBM Corporation 2020

Figure 4-7. Template Folder Structure

The upper screen capture shows the folder structure for a template in Windows Explorer. The lower left screen capture shows a list of templates in Windows Explorer. The lower right screen capture shows the names of the templates that are available in the Application wizard.

Template has the same basic internal configuration as other applications.

- Document Hierarchy (DCO)
- A collection of Rulesets
- Task profiles

*Figure 4-8. Create an application*

The upper screen capture shows the options available to create an application or copy an existing application in Datacap Application wizard. The lower screen capture shows the application name, folder location where the application is added, and the options for the template. You start Datacap Application Wizard in Datacap Studio.

You create a Datacap application whenever there is a need for a new application to process different types of documents. Also, when a capture solution is required for a new department or a document handling procedure is being implemented. You can also copy an existing application to create a new one.

Before you create an application, decide on:

- An application name
- A template or an application for a starting point

Datacap Folder

- It is the root folder for locating the templates that are displayed in the application template list.
- The Datacap folder is normally on the local system but it might be on another server.
- To select a folder on another server, the path is: \\<server name>\Datacap
- The wizard would expect to find a Template folder in this root folder.

Destination

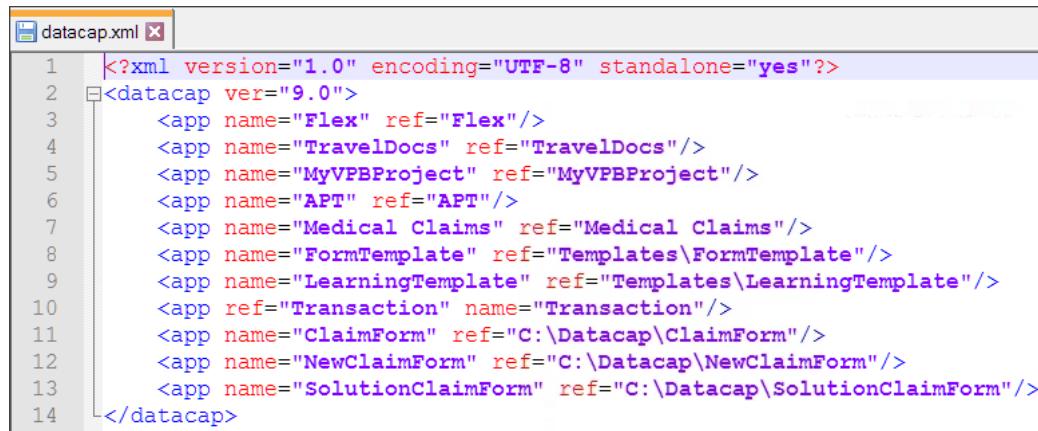
- It is the root folder for placing the application.
- This location might be on the local system but it is more likely to be on the Datacap Server.
- To select a folder on another server, the path is: \\<server name>\Datacap
- The wizard would create the application at: \\<server name>\Datacap\<application name>

Application Template

- Form Template is for known static page formats.
 - You know the types of data you want to capture and where the data is on each image.
 - Examples: Application forms or Beneficiary form
- Learning Template is for more dynamic documents.
 - You know the types of data you want to capture.
 - You do not know where the data is contained on each image. It might be different on each image.
 - Examples: Invoices or Inventories

Datacap.xml

- For every new application, Datacap adds an entry for the new application in the Datacap.xml file



```

1  <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2  <datacap ver="9.0">
3      <app name="Flex" ref="Flex"/>
4      <app name="TravelDocs" ref="TravelDocs"/>
5      <app name="MyVPPProject" ref="MyVPPProject"/>
6      <app name="APT" ref="APT"/>
7      <app name="Medical Claims" ref="Medical Claims"/>
8      <app name="FormTemplate" ref="Templates\FormTemplate"/>
9      <app name="LearningTemplate" ref="Templates\LearningTemplate"/>
10     <app ref="Transaction" name="Transaction"/>
11     <app name="ClaimForm" ref="C:\Datacap\ClaimForm"/>
12     <app name="NewClaimForm" ref="C:\Datacap\NewClaimForm"/>
13     <app name="SolutionClaimForm" ref="C:\Datacap\SolutionClaimForm"/>
14 </datacap>

```

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-9. Datacap.xml

The screen capture shows the Datacap.xml file in Notepad++.

The Datacap.xml file is stored in the C:\Datacap folder on the Datacap server system.

This file maintains a list of applications that are available on the system. This file is called to show the list of applications in the Datacap clients.

For every new application, Datacap adds an entry for the new application in the Datacap.xml file.

The Application wizard does not create the datacap.xml file. It adds an entry for the new application.

4.2. Configure the application in IBM Content Navigator



Figure 4-10. Configure the application in IBM Content Navigator



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-11. IBM Content Navigator administration - create a repository

Datacap Navigator clients are built on the IBM Content Navigator framework. To access the new Datacap applications in the Datacap Navigator client, you need to configure the application in the IBM Content Navigator administration tool.

You use the Datacap Navigator clients to configure tasks and jobs in workflows for the Datacap applications and process batches.

The upper screen capture shows the Repositories tab in the IBM Content Navigator administration tool. On this tab, the action to configure a Datacap application as a repository is highlighted.

The lower screen capture shows the configuration for a new repository. You provide the wTM URI to access the Datacap Server and configure the selected application.



Add a repository to the Datacap admin desktop

The screenshot shows the 'Datacap Admin Desktop' interface with the 'Repositories' tab selected. The 'Selected Repositories' pane lists 'TravelDocs' and 'ClaimForm' (which is highlighted with a red box). The 'Available Repositories' pane lists 'LoanProcess', 'MyVPBProject', and 'Sales'. Navigation arrows (up, down, left, right) are used to move repositories between the two panes.

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-12. Add a repository to the Datacap user desktop

The previous slide shows how to configure a Datacap application as a repository in the IBM Content Navigator administration tool. The repository must be associated with the Datacap Admin Desktop that is Datacap Navigator admin console so that the new application is accessible in the Datacap Navigator clients.

The screen capture shows the Repositories tab for the Datacap Admin Desktop in the IBM Content Navigator administration tool.

- On this tab, you can associate a repository for a Datacap application.
- The Available Repositories pane lists all the repositories that are configured for the desktop.
- For the users to have access, the repositories must be moved to the Selected Repositories pane.



Configure jobs for a workflow in Datacap Admin Console

The screenshot shows the Datacap Admin Console interface. On the left, there's a sidebar with links: Workflows (which is the active tab), Groups, Users, Stations, Shortcuts, Panels, and Redaction. The main area is titled 'Workflow: ClaimForm'. It has tabs for 'General' and 'Jobs', with 'Jobs' being the active one. Below the tabs are buttons for 'New Job', 'Edit', 'Refresh', and 'Delete'. A table lists the jobs:

Name	Description
Claim Form Job	Standard processing of single page images
Navigator Job	Standard processing with Datacap Navigator

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-13. Configure jobs for a workflow in Datacap Admin Console

In the Datacap Navigator Admin Console, you can configure the workflows for the Datacap applications and manage users, groups, and panels.

The Application wizard creates various types of jobs for the application. In the Datacap Navigator Admin Console, you can open the workflow for an application, review the jobs, and delete the ones that you don't need. It is optional to delete the jobs but helps keep the application only with the essential elements. You can also edit the default Jobs to customize to your business needs.

The screen capture shows a list of jobs for an application in the Jobs subtab in the Datacap Navigator Admin Console. On this tab, you can delete the jobs that you don't need, create new jobs, and modify the existing jobs.



Add a task for a job in Datacap Admin Console

Job: Claim Form Job

Name	Description	Mode	Program	Queue By
VScan	Run VScan Rules	Batch creation	rulerunner.exe	None
PageID	Page Identification Rules	Router	rulerunner.exe	None
Profiler	Recognize/Validate w/Rules	Normal	rulerunner.exe	None
Verify	Verify with Rule Validation	Normal	super.aspx	None
Export	Export via Rules	Normal	rulerunner.exe	None

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

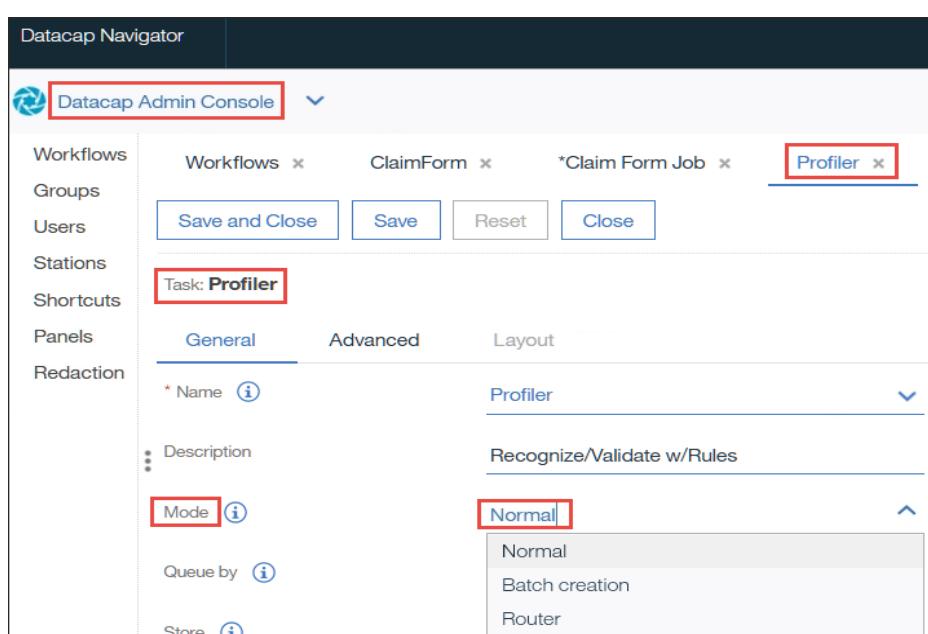
Figure 4-14. Datacap Admin Console - add a task for a job

The Application Wizard creates default tasks for a job in the workflow of an application. In the Datacap Navigator Admin Console, you can review the tasks and add other tasks to customize to your business needs.

The screen capture shows the tasks subtab for a job in the Datacap Navigator Admin Console. On this tab, you can add new and modify the existing tasks. You can also delete the tasks and reorder them.

IBM Training 

Edit a task for a job in Datacap Admin Console



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-15. Datacap Admin Console - edit a task for a job

The Application Wizard creates default task configuration for a job in the workflow of an application. In the Datacap Navigator Admin Console, you can review the tasks and edit the default settings to customize to your business needs.

The screen capture shows the configuration page for Profiler task of a job in the Datacap Navigator Admin Console. On this tab, you can modify the existing settings.

In this example, the value for the Mode field is changed to Normal. If the value is Router, the workflow is routed to a Fixup job.

4.3. Edit the application settings in Datacap Application Manager

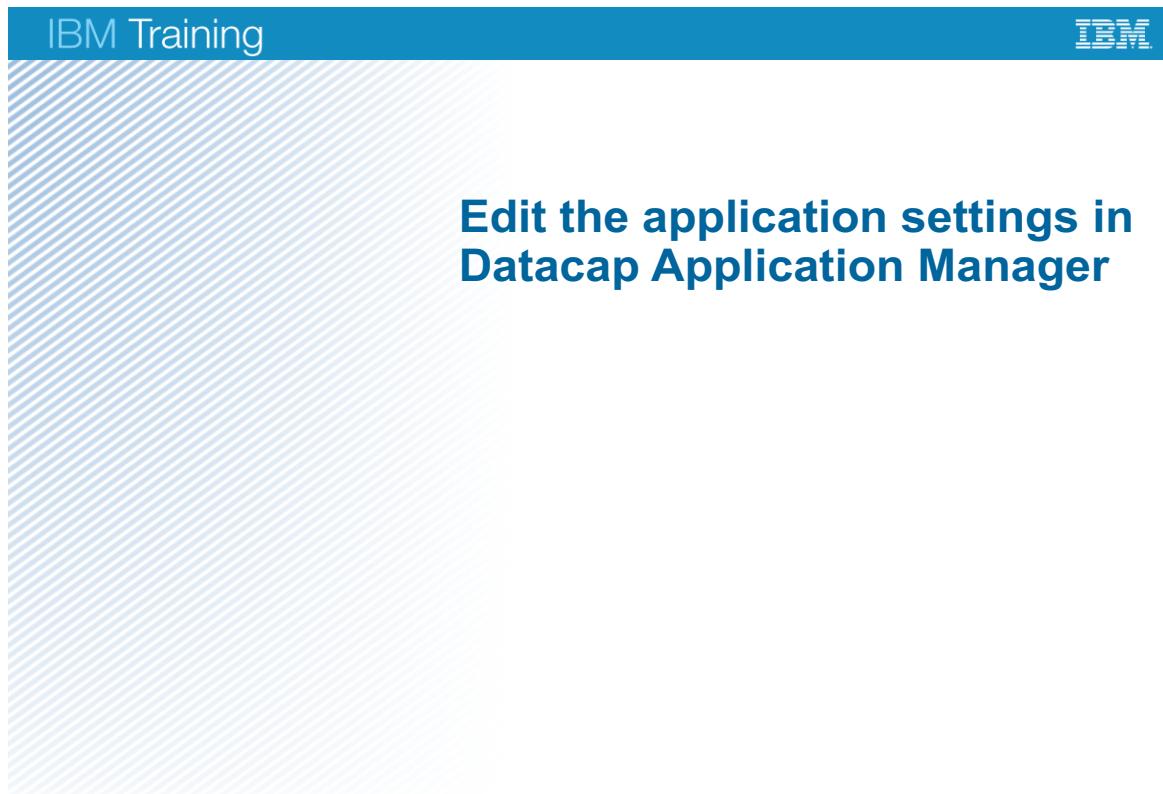


Figure 4-16. Edit the application settings in Datacap Application Manager

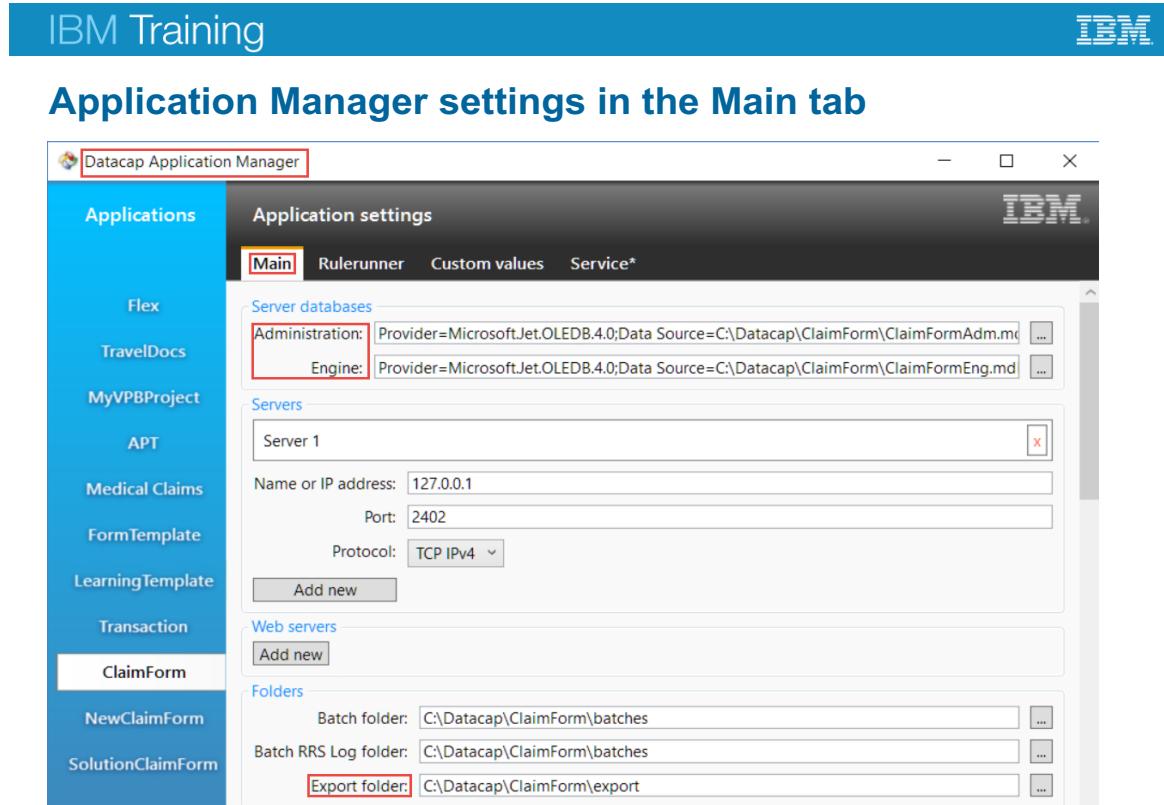


Figure 4-17. Application Manager settings

The screen capture shows the Main tab in Datacap Application Manager.

When you create an application, the wizard saves the default application configuration in the application configuration file. You can view and edit these settings in Datacap Application Manager.

Datacap Application Manager lists all the applications that are registered in Datacap.xml.

The Main tab shows the:

- Settings for Admin and Engine, Fingerprint databases
- Folders for images to be scanned (VScan source folder) and storing the extracted data (Export folder).
- Option to enable FPXML



Application Manager settings in the Rulerunner tab

The screenshot shows the Datacap Application Manager interface. The left sidebar lists applications: Flex, TravelDocs, MyVPBProject, APT, Medical Claims, FormTemplate, LearningTemplate, Transaction, and ClaimForm. The 'ClaimForm' application is selected. The main panel has tabs: Main, Rulerunner (which is highlighted with a red box), Custom values, and Service*. The 'Rulerunner' tab displays task profiles for each task. Task 1: VScan has a Task profile: VScan. Task 2: PageID has a Task profile: PageID. Task 3: Profiler has a Task profile: Profiler. Task 4: Export has a Task profile: Export. There is also an 'Add new' button.

Task	Task profile
Task 1: VScan	VScan
Task 2: PageID	PageID
Task 3: Profiler	Profiler
Task 4: Export	Export

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-18. Application Manager settings

The screen capture shows the Rulerunner tab in Datacap Application Manager. The tab lists the task profiles that you can modify.

In the Datacap Application Manager, for each task, a Task and Task profile pair are listed. For example, two VScan and two PageID.

Task 1: VScan is linked to the database, and Task profile: VScan is linked to the collection.xml.

The tasks that are listed in Datacap Navigator use data from the database and the task profiles that are listed in Datacap Studio uses data from the collection.xml file.

Datacap Application Manager maps the tasks in Datacap Navigator with the task profiles in Datacap Studio.

4.4. Define document hierarchy

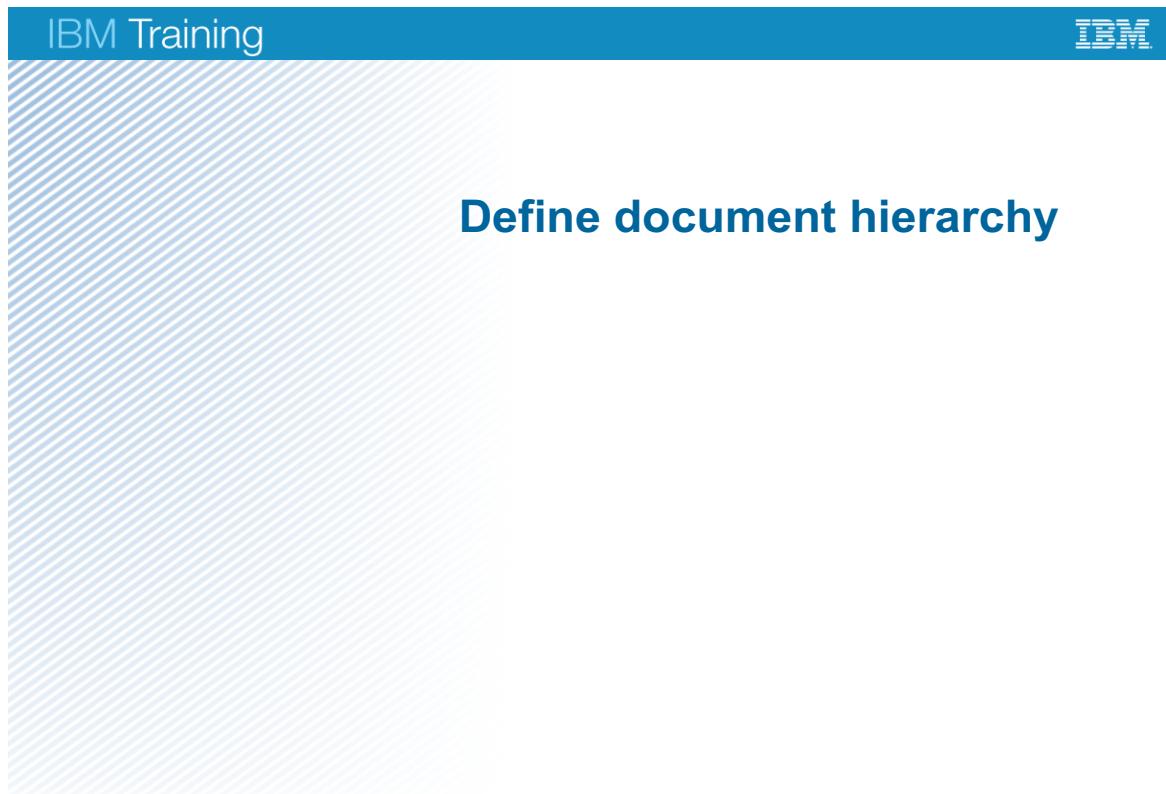


Figure 4-19. Define document hierarchy

Configuring Datacap application in Datacap Studio

- In the Rulemanager tab of Datacap Studio, you define three aspects for the application:
 1. Rulesets
What actions do run?
 2. Task profiles
When do the actions run?
 3. Document hierarchy
Where do the actions run?
- In the Zones tab of Datacap Studio, you define fingerprints and zones
- In the Test tab of Datacap Studio, you test and debug the application

Figure 4-20. Configuring Datacap application in Datacap Studio

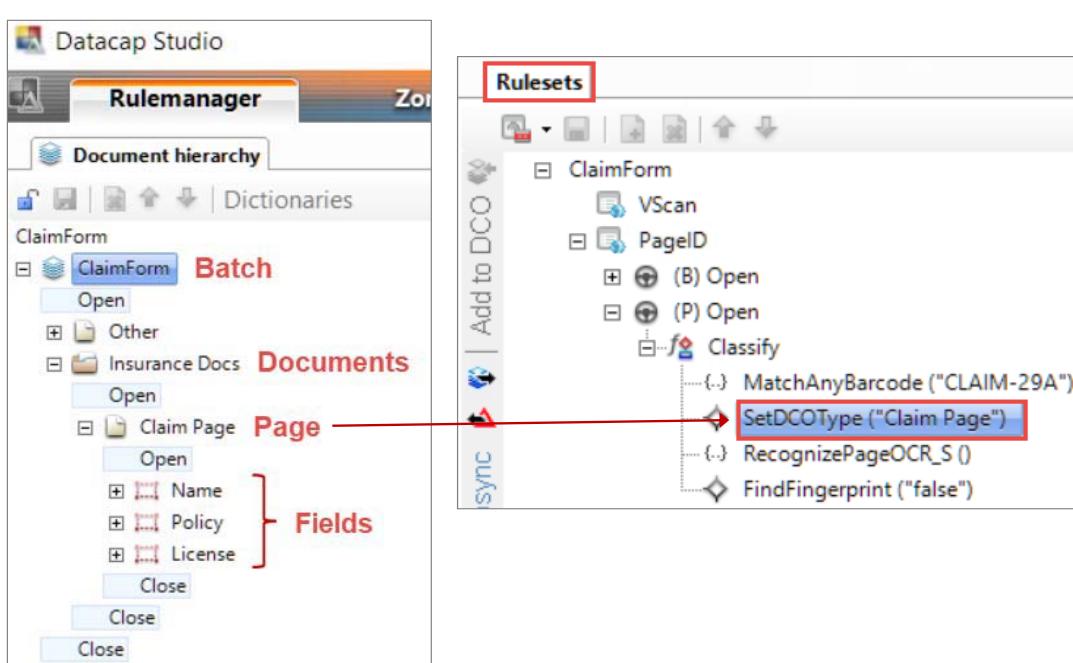
You create an application framework in Datacap application wizard with the Forms template. The framework is used as a starting point to build the application.

The three main panes in the Rulemanager tab of Datacap Studio represent three aspects for the Datacap application workflow: when does it happen (Task profiles), what happens (Rulesets), and where does it happen (Document hierarchy).

IBM Training



Define Document hierarchy objects



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-21. Define Document Hierarchy (DCO) objects

The left screen capture shows the Document Hierarchy (also known as DCO) pane in Datacap Studio Rulemanager tab.

On this pane, you can modify the default DCO (Document, Page, and Field) objects based on your business needs.

In the previous unit, you learned about the application design plan. Before you develop the application, review the documents that are used for scanning and decide the names for the Document, Page, and Field objects. The names for the application, document, page, or field objects must be unique. You receive an error when you try to enter the same name.

The component levels within the hierarchy are Batch, Document, Page, and Field.

- For each application, batch (Example: ClaimForm) is the root level object.
- Each batch contains one or more document types (Example: Insurance Docs)
- Each document type has at least one page type (Example: Claim Page)
- Each page type has a number of fields (Examples: Name, Policy, and License)

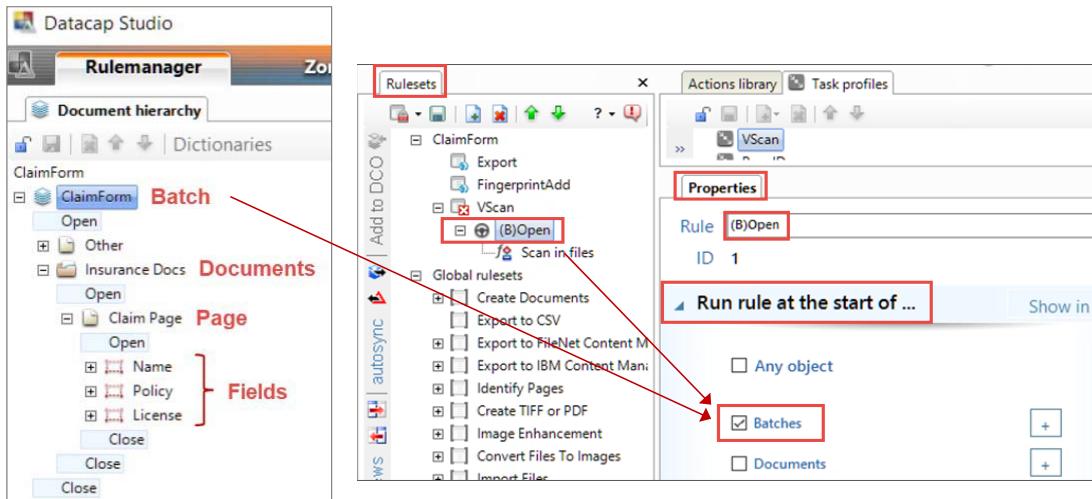
In addition, a page type with the name Other is associated with the batch. The page type is used to temporarily assign a page type to all incoming pages until each page is identified in the application.

Each component has an Open and a Close node in the hierarchy. The Open and Close nodes are used to assign processing rules to the beginning or the end of each component.

You use the DCO objects that are designed in the DCO pane for defining the rulesets and associate rules with the objects.

For example, the right screen capture shows the SetDCOType action in the Page ID ruleset uses the Page type: Claim Page.

Associate rulesets with Document hierarchy objects



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-22. Associate rulesets with Document hierarchy (DCO) objects

The left screen capture shows the Document Hierarchy (also known as DCO) pane in Datacap Studio Rulemanager tab.

The right screen capture shows the rulesets in the Rulesets pane and properties for the (B) Open rule in the Properties pane.

After you define the Document hierarchy (DCO) and configure rules that define what actions to run, you associate the rule with a DCO. This association indicates what actions run on which Datacap object.

The example shows actions in the (B) Open rule is configured to run on batch object in DCO.

4.5. Configure fingerprints and zones

Configure fingerprints and zones

Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-23. Configure fingerprints and zones

Fingerprints

- In Datacap applications, you can use fingerprints for two basic functions:
 1. Page identification to determine whether the incoming page matches a known page
 2. Identifying the field positions for each variant of each known page type
- You can add fingerprints to the fingerprint library from the Datacap Studio Zones tab.
 - Each time that you add a fingerprint, Datacap starts the FingerprintAdd rule set.

Figure 4-24. Fingerprints

Page identification with fingerprints

- Fingerprinting
 - Identifies the current page based on fingerprint match
 - Datacap generates a fingerprint that describes each incoming page
- A fingerprint consists of an image file (.tif) and a recognition file (.cco)
- The CCO contains the location of all words and lines
- Example: Compare dark and light zones to identify an incoming page

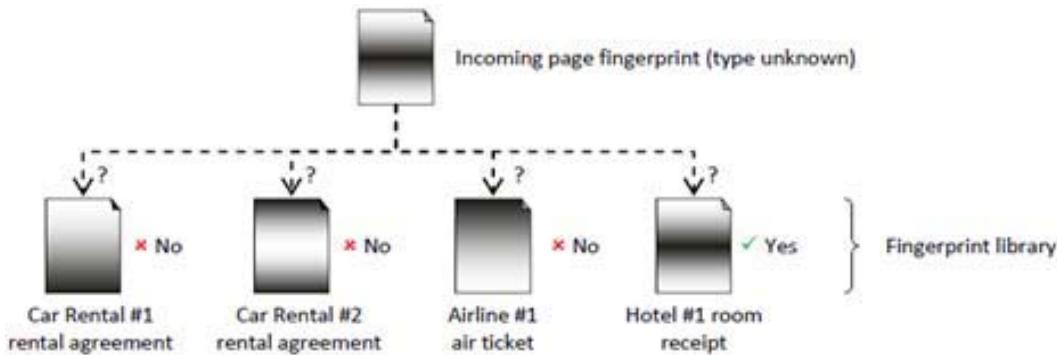


Figure 4-25. Page identification – Fingerprints

How does the Fingerprinting technique work?

- The fingerprint can include information about the relative densities of different regions of the page or the location of text on the page.
- Datacap compares an incoming page fingerprint to existing ones.
- If it matches an existing fingerprint, it is safe to assume that the incoming page is of the same class as the existing one.
 - The offset required to give the best match is also captured.
- If a match does not occur, it creates a fingerprint.
- Technique is adapted for structured and semi-structured documents with a fairly constant layout.

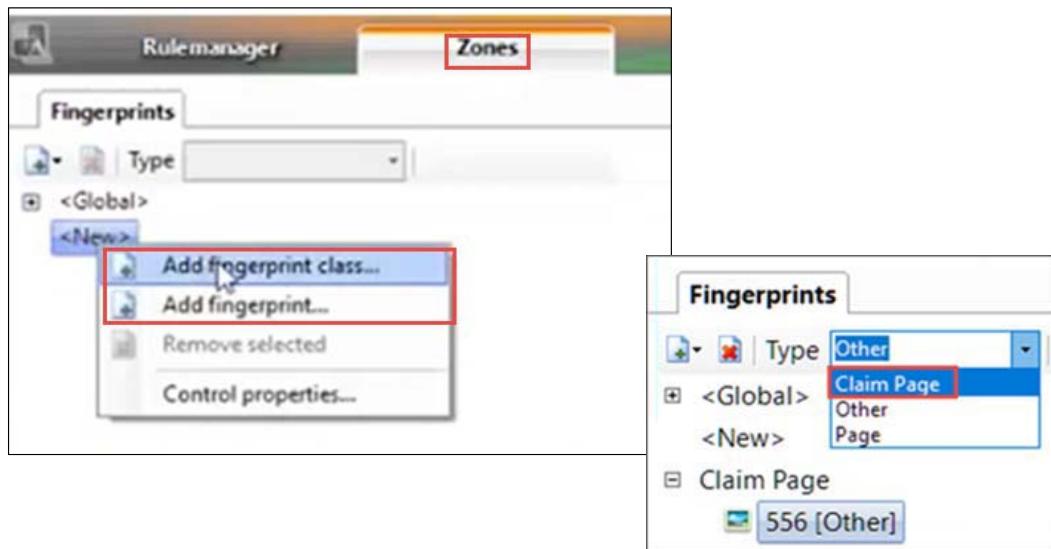
Example of page identification

In the example that is shown on the diagram, the incoming page matches the Hotel #1 room receipt.

Datacap assigns it the type “Room_Receipt”, and records the ID of the matching fingerprint in the runtime batch hierarchy.

Configure fingerprints in the Zones tab

1. Add a Fingerprint class
2. Add a Fingerprint - select an image and assign a page type



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-26. Configure Fingerprints in the Zones tab

The screen capture shows the Zones tab. In the upper right screen capture, the Add fingerprint class and Add fingerprint actions are highlighted. The lower left screen capture shows the selection of a page type for the fingerprint.

The Add fingerprint class action creates a fingerprint class. You can add fingerprints to this class by using the Add fingerprint action.

When you add a fingerprint, it prompts you select an image file. This step runs the image file, cleans the image, and creates the Fingerprint. You can assign a type to the new fingerprint (a page type that you defined in DCO).

Fingerprint classes:

- Are helpful to differentiate the forms that you receive from various sources.
- Can be used to group fingerprints based on the source of the form.

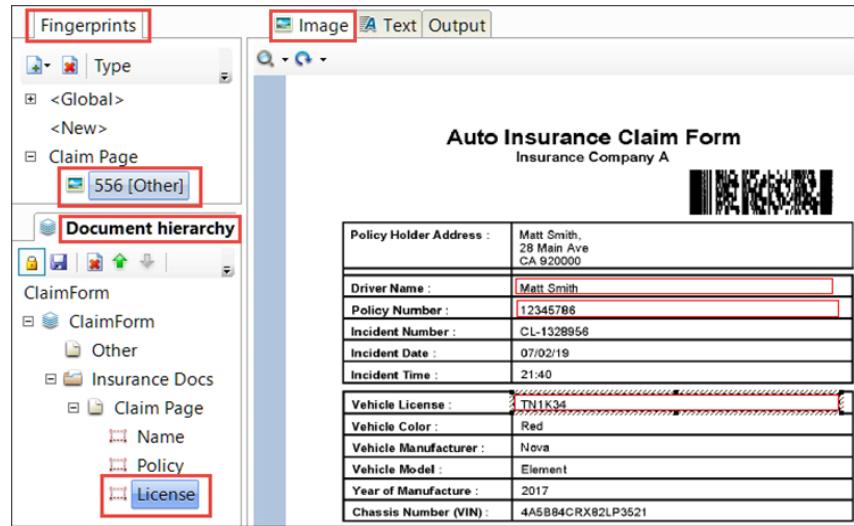
The page type names in the list come from the <application name>.xml file in the C:\Datacap\<application name>\dco_<application name>\rules folder.

Example: C:\Datacap\ClaimForm\dco_ClaimForm\rules\ClaimForm.xml

The fingerprints that are available for the application is listed in the Datacap/ClaimForm/fingerprint folder.

IBM Training

Configure zones



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-27. Configure Zones

The screen capture shows the Datacap Studio Zones tab. In the Fingerprints pane, a fingerprint is selected (556[Other]), and in the Document Hierarchy pane, a field is selected. On the Image tab, the image that is associated with the fingerprint (556[Other]) is selected and zones are drawn for the selected field.

- Fingerprints are defined to enable the identification of images by the position of data items on a page.
- When an image is identified by matching it to a fingerprint, you can extract the data from the fields that are defined as zones in the fingerprint.
- Locating data on an image with the zonal information that is stored in the fingerprint is more efficient than other methods of locating data like doing keyword searches.



Set OCR filters for the fields

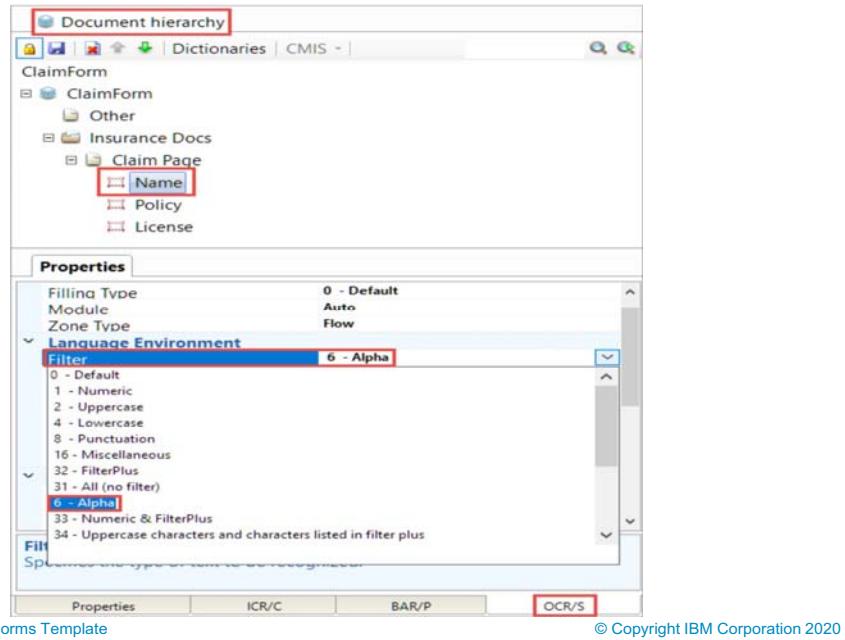


Figure 4-28. Set OCR filters for the fields

The OCR filters specify the type of text to be recognized.

The screen capture shows the Datacap Studio Zones tab. A field (Name) is selected in the Document Hierarchy pane and in the Properties tab, 6-Alpha is selected for the Filter field. This selection specifies that the alphabetic characters are recognized for this field.

Unit summary

- Describe how to create a Datacap application with Forms Template
- Learn the application configuration in IBM Content Navigator
- Identify the application settings in Datacap Application Manager
- Describe document hierarchy
- Learn about the fingerprints and zones configuration

Figure 4-29. Unit summary

Review questions

1. True or False: Application Wizard generates a basic application framework with the required directories, databases, and files.
2. True or False: For every new application that you create, you must manually add an entry for the application in the Datacap.xml file.
3. In the IBM Content Navigator administration tool, you configure your Datacap application to do which of the following tasks in the Datacap Navigator clients? (Choose two)
 - A. Configure tasks and jobs in workflows
 - B. Process batches
 - C. Define document hierarchy
 - D. Create rulesets



Figure 4-30. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. True or False: Application Wizard generates a basic application framework with the required directories, databases, and files.
The answer is True.
2. True or False: For every new application that you create, you must manually add an entry for the application in the Datacap.xml file.
The answer is False.
3. In the IBM Content Navigator administration tool, you configure your Datacap application to do which of the following tasks in the Datacap Navigator clients? (Choose two)
 - A. Configure tasks and jobs in workflows
 - B. Process batches
 - C. Define document hierarchy
 - D. Create rulesets
 - E. The answer is A and B.



Building a Datacap application with Forms Template

© Copyright IBM Corporation 2020

Figure 4-31. Review answers

2. For every new application, Datacap automatically adds an entry for the new application in the Datacap.xml file.
3. You define document hierarchy and create rulesets in Datacap Studio.

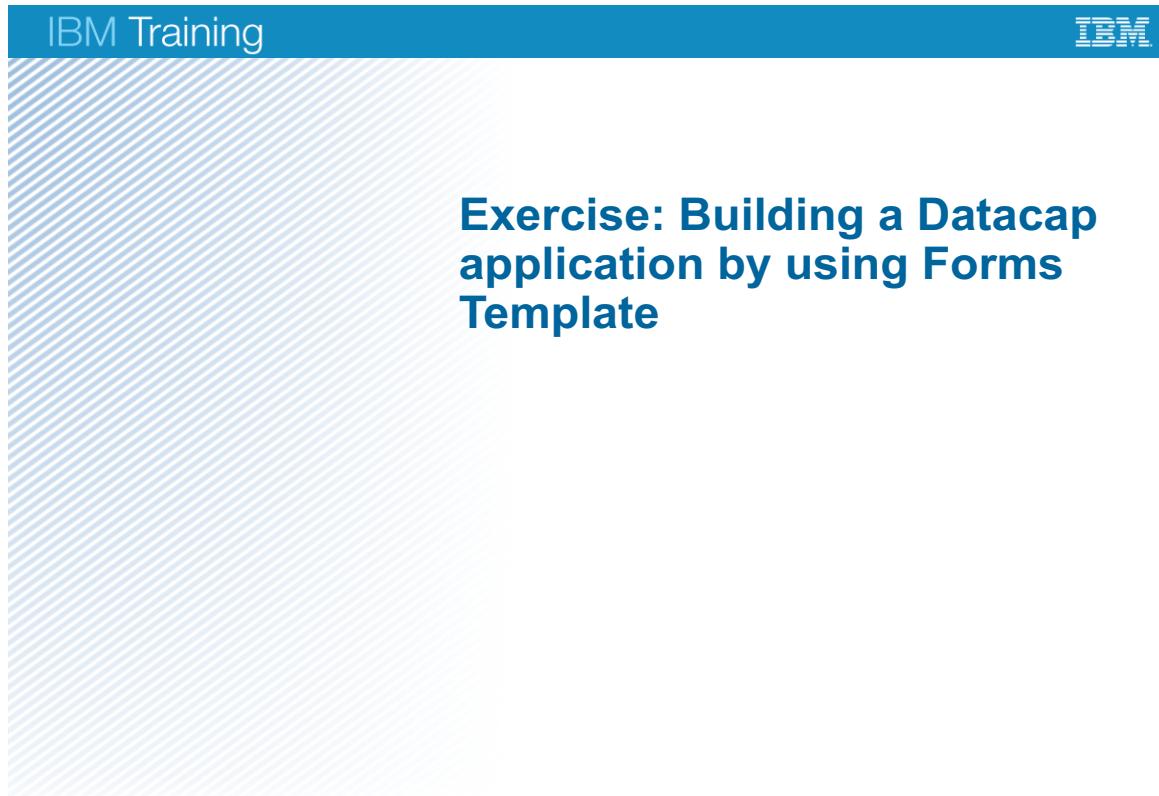


Figure 4-32. Exercise: Building a Datacap application by using Forms Template

Exercise objectives

- Create an application with Forms Template
- Configure the application in IBM Content Navigator
- Edit the application settings in Datacap Application Manager
- Define document hierarchy
- Configure fingerprints and zones



Figure 4-33. Exercise objectives

Unit 5. Configuring rulesets for a Datacap application

Estimated time

00:45

Overview

In this unit, you learn how to add rulesets to a Datacap application, associate them with task profiles, and test the application.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Learn how to add rulesets to a Datacap application
- Describe the role of task profiles
- Learn how to test the application in Datacap Studio

Figure 5-1. Unit objectives

Topics

- Add rulesets
- Configure task profiles
- Test the application

Figure 5-2. Topics

5.1. Add rulesets

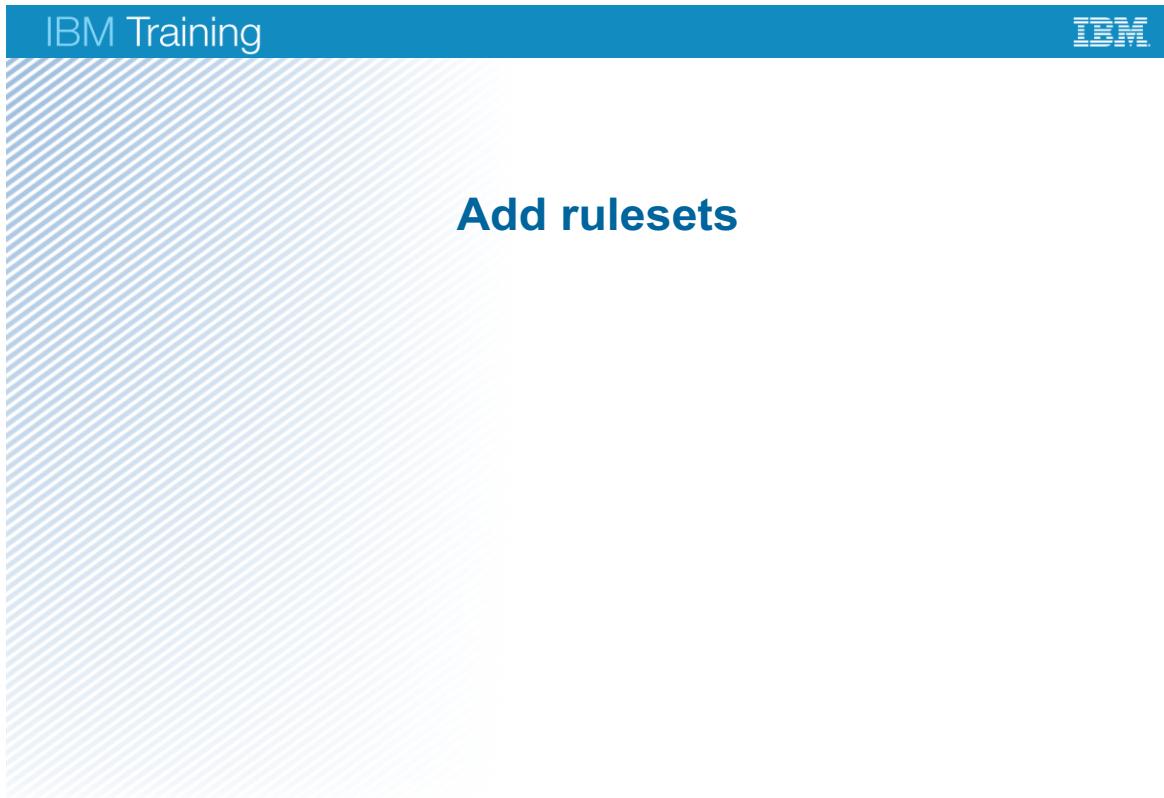
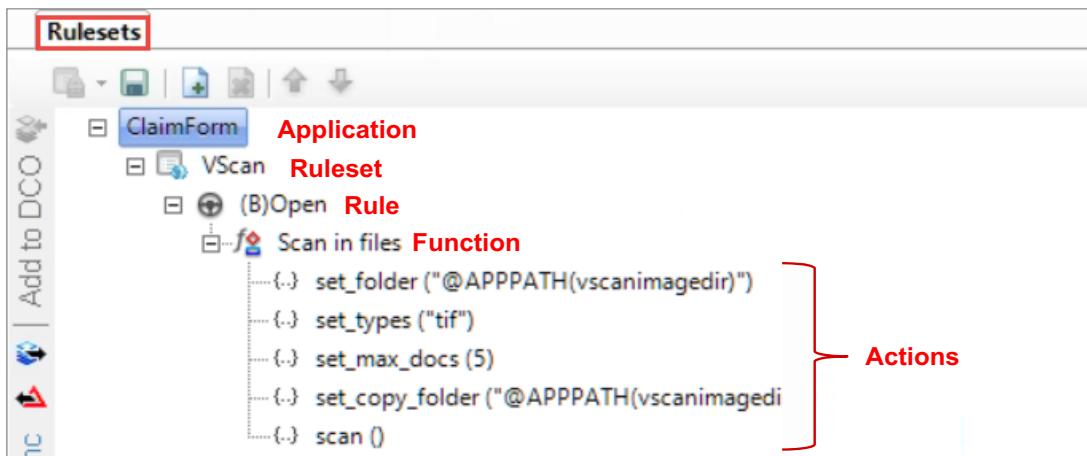


Figure 5-3. Add rulesets

Rulesets, rules, functions, and actions



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-4. Rulesets, rules, functions, and actions

The screen capture shows a sample ruleset and its components (Rules, Functions, and Actions) in the Rulesets pane of the Datacap Studio Rulemanager tab.

- Each Datacap application contains many rulesets
Example: VScan
- Each ruleset contains one or more rules
Example: (B) Open
- Each rule contains one or more functions
Example: Scan in files
- Each function contains one or more actions
Example: set_types("tif")

Rules

As you see in the previous slide, rules are assigned to process specific DCO objects in the document hierarchy (for example, to analyze each page and identify its type). An ordered set of functions (and actions) that are defined in the rule process an object. The rule mapping is discussed in the following slides.

Functions

A function is an ordered list of actions that run until one of the actions returns a value of False.

Rules are made up of functions that run in the listed order until one of the functions completes or all functions failed.

Actions

Actions are the fundamental building blocks that are used to define rules.

- They always return True or False to the Rulerunner service and might run other types of actions.
- Built with .Net or VB script
- A group of actions forms a function
- Actions library provides prebuilt actions

Naming convention of the rules:

Observe that the rules are labeled as (B) Open, (P) Open

The screenshot shows two main panes. The left pane, titled 'Rulesets', lists a single item: 'ClaimForm' under 'Add to DCO'. Inside 'ClaimForm' are several actions: VScan, PagelD, CreateDocs&Fields, Recognize, Validate, Export, and FingerprintAdd. The right pane, titled 'Actions library' and 'Task profiles', shows a tree view of task profiles. The 'Task profiles' tab is selected. A vertical list on the left says 'Add ruleset to profile'. The tree structure shows: VScan (under ClaimForm), PagelD, Profiler, Verify, Export, and FingerprintAdd. Arrows point from each action in the 'Rulesets' pane to its corresponding node in the 'Task profiles' tree. To the right of the tree is a vertical workflow diagram with rounded rectangles and downward arrows: VScan → PagelD → Profiler → Verify → Export.

Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-5. List of rulesets and task profiles that are used for the application

The leftmost screen capture shows an example list of rulesets in the Rulesets pane in Datacap Studio. These rulesets are used for this unit.

The rightmost diagram shows sample workflow steps. Compare them with the Task profiles in the middle screen capture.

The rulesets are associated with task profiles. Compare the rulesets in the Rulesets pane with the rulesets that are on the Task profiles tab. Each task runs the rulesets in the order it is listed in the task profile.

After you configure rulesets that define what actions to run, you associate the rulesets with task profile. This association indicates what actions run when (which step in the workflow).

To keep the rulesets organized, you can order them in the Rulesets pane to match with the order in Task profiles (although it is not required).

FingerprintAdd

This ruleset is specific name. Datacap Studio Zones tab uses it. It is not tied to the tasks in the workflow, but Datacap recognizes it.

Removing the rulesets that you don't want.

- You delete the rulesets associations in the Task profile pane and also delete them from the Rulesets tab.
- The rulesets are stored in the C:/Datacap/ClaimForm/dco_ClaimForm/rules folder and you can also delete them in Windows Explorer.

Collection.xml

The file is stored in the C:\Datacap\<app_name>\dco_<app_name>\rules folder. This file maintains the list of the rulesets and task profiles that are used in the application.

Example folder: C:\Datacap\ClaimForm\dco_ClaimForm\rules

Toolbar in Datacap Studio Rulesets pane

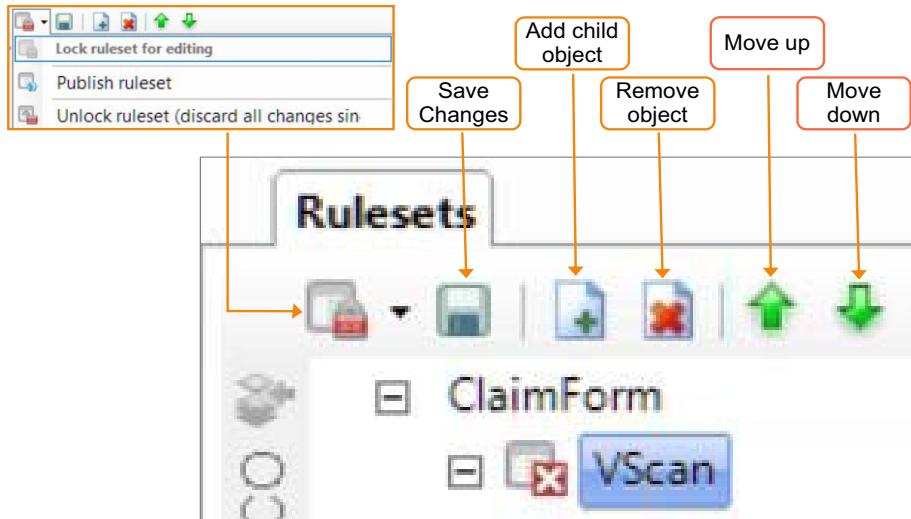


Figure 5-6. Toolbar in Datacap Studio Rulesets pane

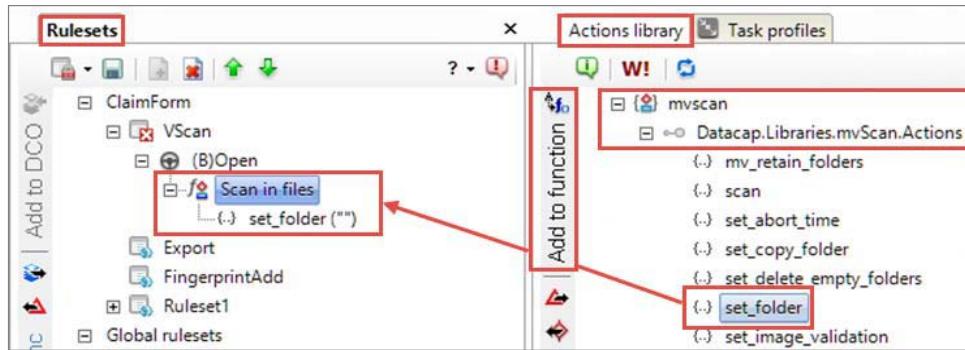
The screen capture shows the toolbar for the Rulesets pane. On this pane, you can edit the default rulesets that are created by the application wizard and add new rulesets.

Actions on the toolbar:

- Lock/Unlock ruleset for editing - Lock a ruleset for editing, publish a ruleset, or unlock a ruleset to discard the changes
- Save the changes - Save the changes periodically when you configure the rulesets
- Add child object -Based on the object that you select, you can add the child object for that object
- Select the application and add a ruleset
- Select a ruleset and add a rule
- Select a rule and add a function
- Remove object -Based on the object that you select, you can delete that object (Ruleset, rule, function, or action)
- Move up or Move down - rearrange the objects (Ruleset, rule, function, or action)



Add an action to a function



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-7. Add an action to a function

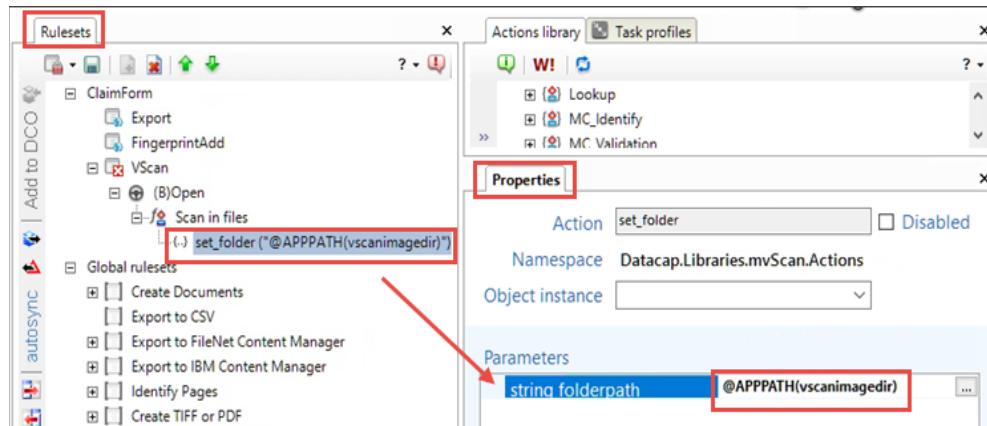
The screen capture shows how to add an action from the actions library to a function in a rule.

In the Rulesets pane, to configure a ruleset, add the ruleset, rule, and function objects. Then, add actions from the actions library to the function.

In the example, VScan ruleset, (B) Open rule, Scan in files function are defined. The Set_folder action from the mvscan library is added.

IBM Training IBM

Set parameters to an action



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

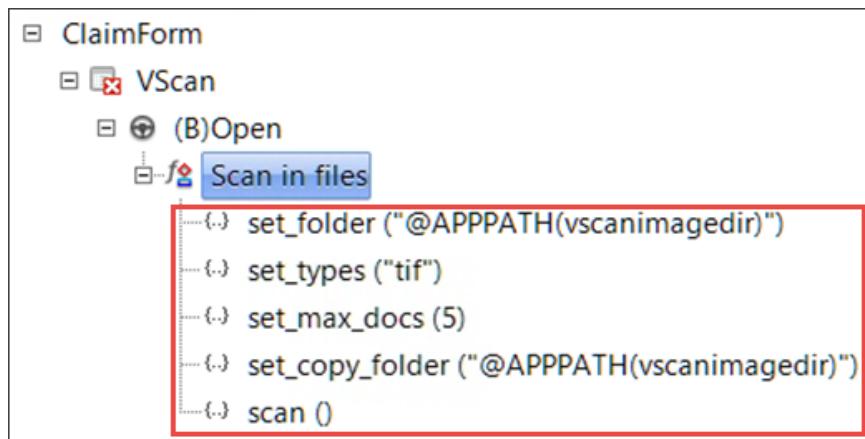
Figure 5-8. Set parameters to an action

The screen capture shows how to set parameters to an action in Datacap Studio.

In the Rulesets pane, after you add an action from the actions library to the function, set the parameters in the properties pane.

IBM Training IBM

VScan ruleset



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-9. VScan ruleset

For scanning the images, you add actions to the VScan ruleset.

This example contains the VScan ruleset, (B) Open rule, and Scan in files function.

The rule name (B) Open indicates that this rule is run at the Batch level. The naming convention helps identify at what level the rule is run.

In a ruleset, due to the order of operations, you can have only one rule at batch level. This naming convention helps to prevent adding multiple rules that run at the same level.

Naming conventions

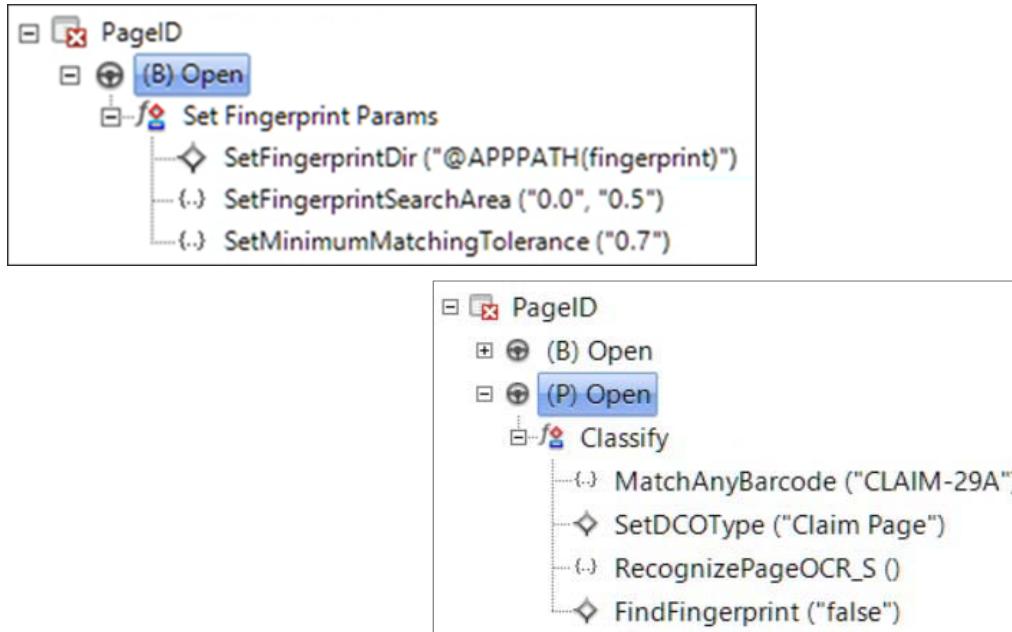
- A Rule is named based on what level it runs (Example: (B) Open, (P) Open, and (F) Open)
- A Function is named based on what it does

The actions in the "Scan in files" rule sets the following values:

- A folder from where images are scanned for the batch
- The type of files
- Maximum number of the documents that can be scanned per batch

The images folder is defined in Datacap Application Manager.

Page ID ruleset



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-10. Page ID ruleset

The screen captures show the Page ID ruleset.

The upper left screen capture shows the (B) Open rule and it is run at the batch level. The actions define Fingerprint Settings. It is best to run the rule at batch level because it is most efficient to load the settings one time per batch. You can run it at the page level, but it is inefficient to load the settings at every page.

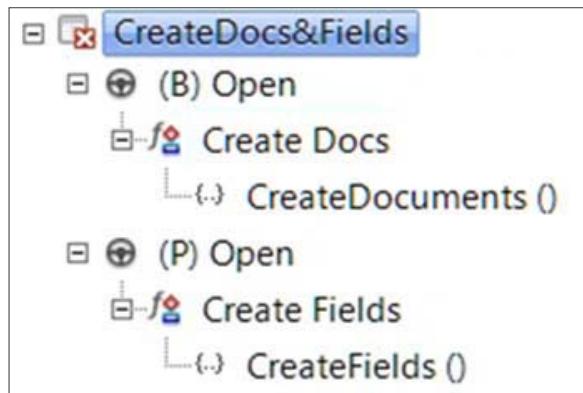
The lower right screen capture shows the (P) Open rule and it is run at the page level. The actions classify the image.

- For the MatchAnyBarcode action parameter, provide the Barcode value from the image. For example, CLAIM-29A. Don't include the asterisk that is shown on the form. The asterisk is part of the Barcode coding. The value in between asterisk is what the Barcode generates.

The MatchAnyBarcode action searches all the barcodes on the current page and checks if any one matches the value you enter as a parameter.

- The SetDCOType action assigns a value to the Type property of the current object (Page) of the Document Hierarchy.
- For the FindFingerprint action, you can set true or false. If scanned image does not match the existing fingerprint, you have options to create more new fingerprints. You set false when you don't want to create more fingerprints.

CreateDocs&Fields ruleset



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-11. *CreateDocs&Fields ruleset*

The screen capture shows the CreateDocs&Fields ruleset. The actions in the CreateDocs&Fields ruleset create document, page, and fields in DCO.

The CreateDocuments action runs at the Batch level and the CreateFields action is run at the page level.

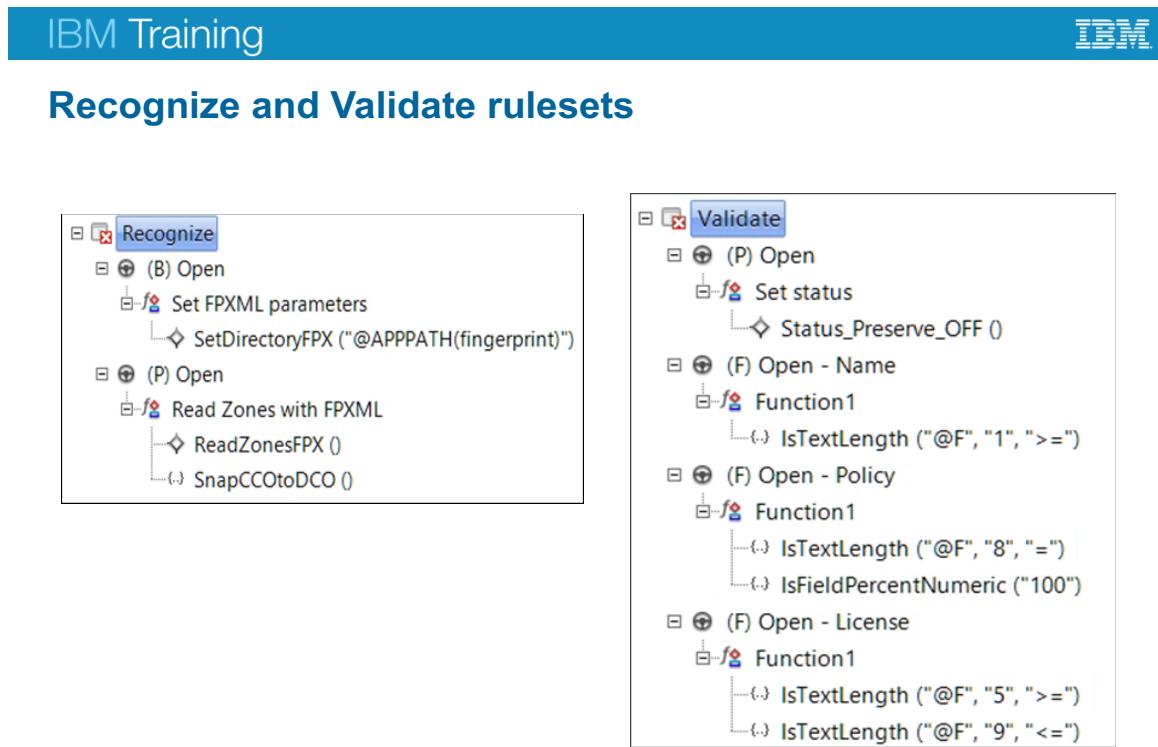


Figure 5-12. Recognize and Validate rulesets

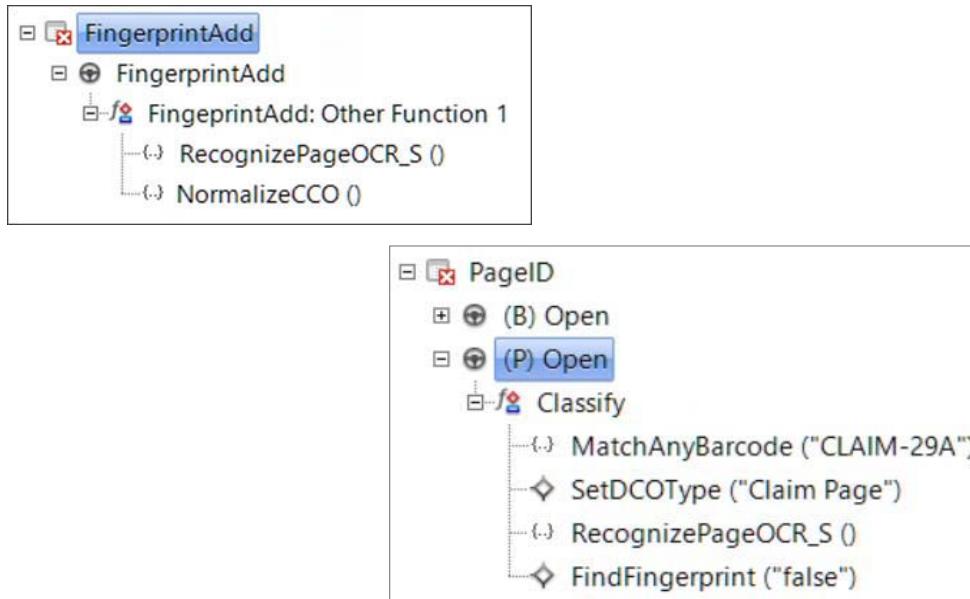
The left screen capture shows the Recognize ruleset:

- You set the directory for FPXML at the batch level and configure the Read Zones action at the page level.
 - The SnapCCOtoDCO() action maps the page recognition results (field values) from CCO file into DCO fields based on the zone position of the fields.

The left screen capture shows the Validate ruleset:

- The Status_Preserve_OFF action sends the verification step to human verification and it is run at the page level.
 - The actions at the Field level validate the fields to check whether the field has value and the data type is correct.

FingerprintAdd ruleset



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-13. *FingerprintAdd ruleset*

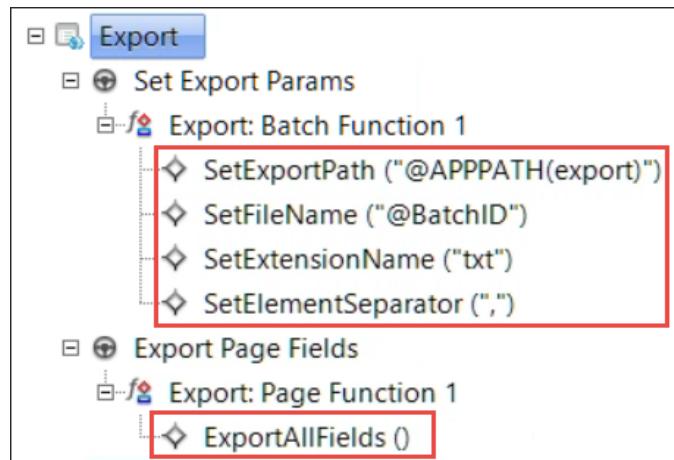
The OCR engine that you used for the PageID ruleset to create the Fingerprint (CCO file) must match with the OCR engine that is used in the FingerprintAdd ruleset. In the exercise, you edit the ruleset to match them.

The screen captures show the FingerprintAdd ruleset on the upper left and the PageID ruleset on the lower right. Notice that both rulesets contain the same OCR action that creates the Fingerprint file: RecognizePageOCR_S() action from the ocr_sr action library.

This ruleset is added to the task profile tab, but not tied to the task in the workflow. Datacap recognizes this ruleset. In the Zones tab, each time you add a fingerprint, Datacap starts the FingerprintAdd rule set.

Export the data

- What does Export mean?
 - Determine where the final capture data is placed and what form it is in when it is exported
 - Export data to a text file, an XML file, and other business processes



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-14. Export the data

The screen capture shows the Export ruleset for exporting the data to a text file. This ruleset is automatically added by the wizard when you create an application.

- The actions in the first (batch level) rule of the Export ruleset sets the following parameters:
 - Folder to export file
 - File name
 - File extension
 - A separator such as comma
- The actions in the second (Page level) rule of the Export ruleset exports all the fields.

Export to a text file

The Export step is the last step of the capture process. When Capture processing is complete, you can set up Datacap to export data to a text file, an XML file, a database, a document management system, or a custom business process.

The default output format is a text file and the default export location is the exports folder for the application. The default export folder is defined in the Datacap Application Manager by the Export folder parameter.

5.2. Configure Task profiles

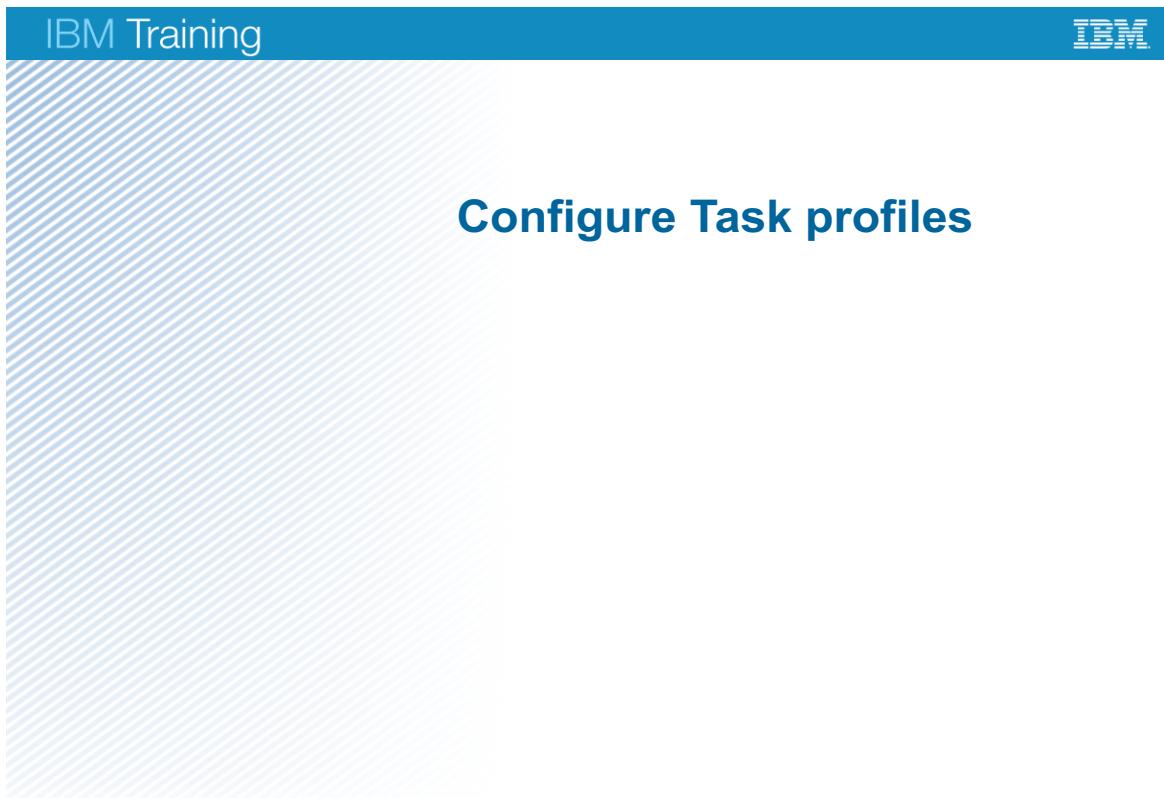
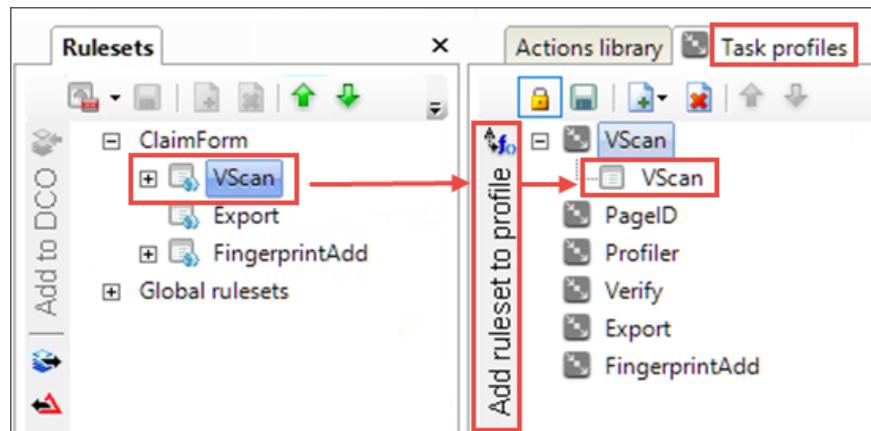


Figure 5-15. Configure Task profiles

Add a ruleset to a task profile



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-16. Add a ruleset to a task profile

Task profiles define in what order, the rulesets are run for a task. You must associate a ruleset with a task profile.

The screen capture shows how to add a ruleset to a task profile in the Task profiles tab. In Rulesets tab, select the ruleset, in the Task profiles tab, select the task profile, and click the Add ruleset to profile button.

Recall that in the Rulemanager tab of Datacap Studio, you define three aspects for the application:

a. Rulesets

What actions do you run?

b. Task profiles

When do you run the actions?

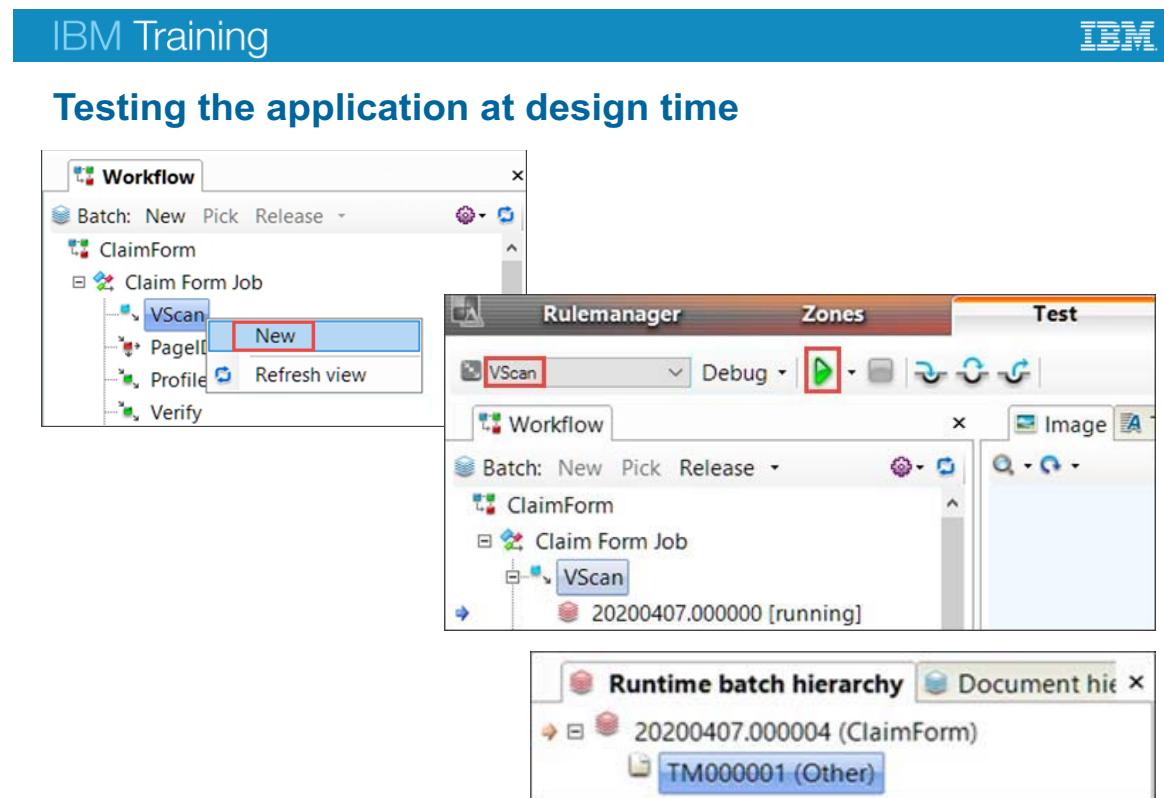
c. Document hierarchy

Where do you run the actions?

5.3. Test the application



Figure 5-17. Test the application



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-18. Testing the application at design time

You can test your application at design time within Datacap Studio.

The screen captures show the Test view in Datacap Studio.

- The upper left screen capture shows how to start a batch at the first task (VScan) in the Workflow pane.
- The middle screen capture shows that VScan task is completed and a batch item is shown under the VScan node.
- The lower right screen capture shows that the scanned page is listed in the Runtime batch hierarchy. At this step, the page type is shown as Other.

You can run all the tasks in a workflow in Datacap Studio Test tab. The following slides show each task.

Testing the application - Advance the batch

Advance the batch to the next task in the workflow.

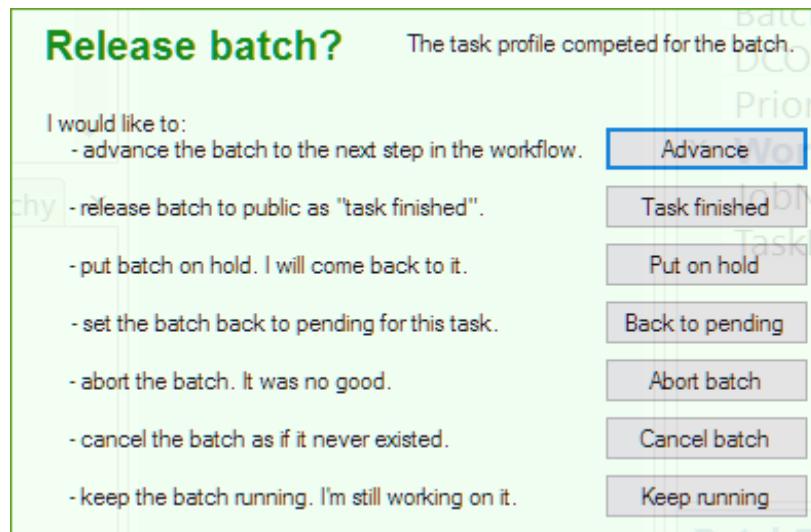
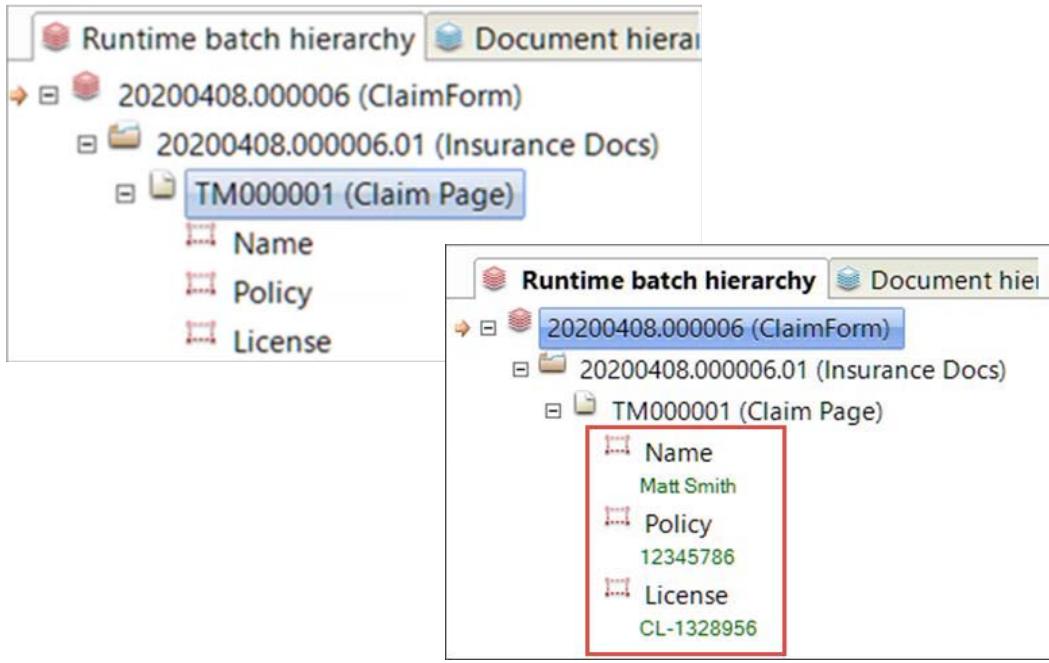


Figure 5-19. Testing the application - Advance the batch

The screen capture shows the Release batch dialog box that is shown when a task is completed.

You can advance the batch to the next task in the workflow. You have other options, for example, to cancel the batch, to put it on hold, or change it to the pending state.

Testing the application - Page ID and Profiler tasks



Configuring rulesets for a Datacap application

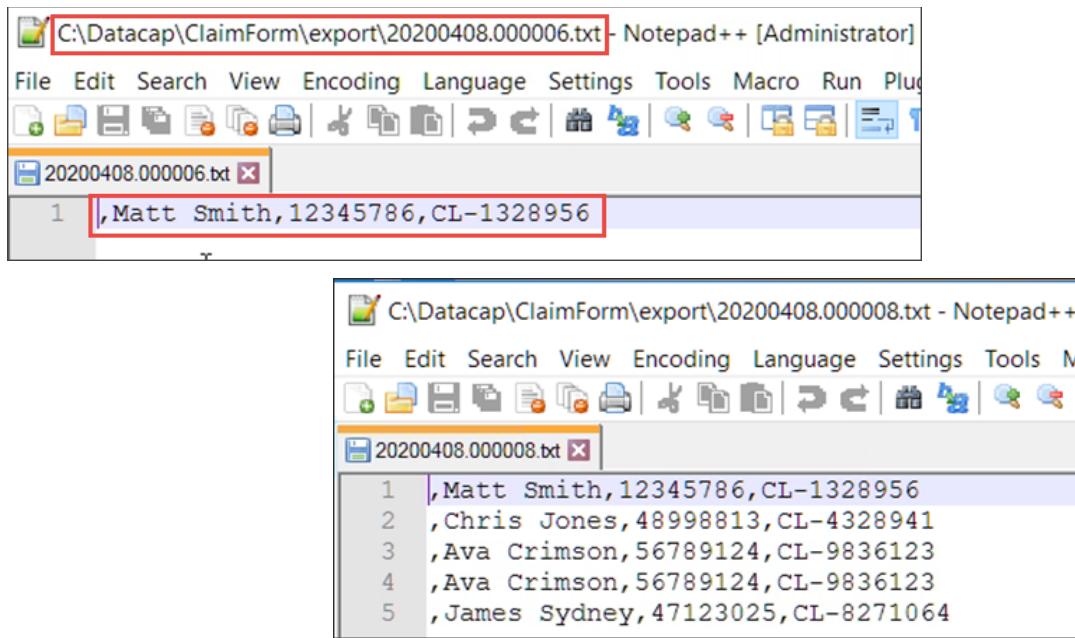
© Copyright IBM Corporation 2020

Figure 5-20. Testing the application - Page ID and Profiler tasks

The screen captures show the batch results in the Runtime batch hierarchy pane of the Datacap Studio Test view.

- The upper left screen capture shows the results for the PageID task. The Page (Claim Page) and Document (Insurance Docs) types are identified. The fields are created (Name, Policy, and License)
- The lower right screen shows the results for the Profiler task. The extracted values are shown in DCO fields.

Testing the application - Export task



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-21. Testing the application - Export task

The screen captures show the export text files in Notepad++. The file is created after the Export task is completed for the batch.

- The upper screen capture shows the export text file with the field values for one scanned image in a batch.
- The lower screen shows the export text file with the field values for five scanned images in a batch.

Unit summary

- Learn how to add rulesets to a Datacap application
- Describe the role of task profiles
- Learn how to test the application in Datacap Studio

Figure 5-22. Unit summary

Review questions

1. Fingerprint Settings rule sets all the parameters for using the FindFingerprint action. At what level do you run the rule?
 - A. Batch
 - B. Document
 - C. Page
 - D. Field
2. At what level do you run the FindFingerprint action?
 - A. Batch
 - B. Document
 - C. Page
 - D. Field
3. If you use fingerprinting and zones to capture data, which of the following actions do you use to extract the data?
 - A. Regexfind
 - B. ReadZones or Readzones FPXML
 - C. Locate
 - D. Dictionary



Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-23. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

- Fingerprint Settings rule sets all the parameters for using the FindFingerprint action. At what level do you run the rule?

- A. Batch
- B. Document
- C. Page
- D. Field

The answer is A.

- At what level do you run the FindFingerprint action?

- A. Batch
- B. Document
- C. Page
- D. Field

The answer is C.



Figure 5-24. Review answers

- Fingerprint Settings rule sets all the parameters for using FindFingerprint action. At what level does the rule run?

Answer – Batch – level. It is the best answer because it is most efficient to load the settings one time per batch.

Answer – Page – level. The action can run at a page level, but it is inefficient to load the settings at every page.

Review answers

3. If you use fingerprinting and zones to capture data, which of the following actions do you use to extract the data?
- A. Regexfind
 - B. ReadZones or Readzones FPXML
 - C. Locate
 - D. Dictionary
- The answer is B.



Figure 5-25. Review answers

Exercise: Configuring rulesets for a Datacap application

Configuring rulesets for a Datacap application

© Copyright IBM Corporation 2020

Figure 5-26. Exercise: Configuring rulesets for a Datacap application

Exercise objectives

- Create rulesets
- Assign the rulesets to task profiles
- Test the application



Figure 5-27. Exercise objectives

Unit 6. Troubleshooting a Datacap application

Estimated time

00:30

Overview

In this unit, you learn how to examine the logs files and troubleshoot the applications in the Datacap Studio.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacapoc.doc/datacap_9.1.7.htm

Unit objectives

- Examine the logs files for Datacap applications
- Learn how to troubleshoot the applications in the Datacap Studio

Figure 6-1. Unit objectives

Topics

- Examine the logs files for debugging
- Troubleshoot the applications in the Datacap Studio Test tab

Figure 6-2. Topics

6.1. Examine the logs files for debugging

Examine the logs files for debugging

Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-3. Examine the logs files for debugging

Logs files that are used in debugging

- As Rulerunner runs each action, it writes detailed logging information to a Rulerunner Service (RRS) log file.
- The RRS logs:
 - Provide detailed information about each action as it is run
 - Are created in the application batches folder

C:\Datacap\ClaimForm\batches\20200408.000007		
Name	Date modified	Type
export_rrs.log	4/8/2020 5:35 AM	Text Document
pageid_rrs.log	4/8/2020 5:35 AM	Text Document
profiler_rrs.log	4/8/2020 5:35 AM	Text Document
verify_rrs.log	4/8/2020 5:35 AM	Text Document
vscan_rrs.log	4/8/2020 5:35 AM	Text Document

Figure 6-4. Logs files that are used in debugging

The screen capture shows the following runtime Rulerunner Service (RRS) logs in Windows Explorer:

- vscan_rrs.log
- pageid_rrs.log
- profiler_rrs.log
- verify_rrs.log
- export_rrs.log

The log files are created in the application batches folder: C:\Datacap\<application name>\batches\<batch name>

Logging is enabled by default for all tasks except VScan and tasks that you start from Datacap Studio.

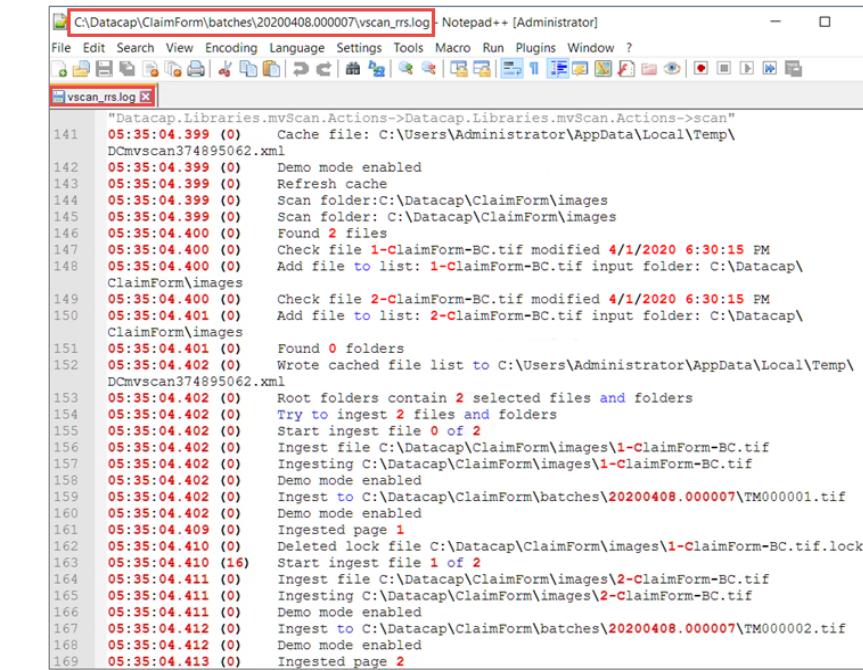
Documentation on Datacap logs:

Search for Datacap logs information by entering the following search string in your web browser:
<https://www.ibm.com/search?lang=en&cc=us&q=datacap%20v9.1.7%20logging>

<https://www.ibm.com/support/pages/node/873182>



Sample log file - task completed successfully



Troubleshooting a Datacap application © Copyright IBM Corporation 2020

Figure 6-5. Sample log file - task completed successfully

The screen capture shows a sample vscan_rrs.log file in Notepad++. The text indicates that the task completed successfully.

In Datacap Studio, rulesets are associated with a task profile. Each ruleset calls actions. You can compare the action names in the log file.

The scan() action is defined in the VScan ruleset and is run for the VScan task.

Observe the following information on the log file:

- In line 140 at the top, the scan action is called.
- In line 144, a previous action (set_folder) is correctly identified the folder that you defined in the application where the images are stored for scanning.
- In the following lines, scan finds the two images files that are available for scanning and ingests the file.
- Line 161 shows that the first file is ingested and line 169 shows that the second file is ingested. It indicates that the VScan is successful.

Sample log file - task was not completed successfully

```

C:\Datacap\ClaimForm\batches\20200408.000007\vscan_rrs.log - Notepad++ [Administrator]
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
vscan_rrs.log

"Datacap.Libraries.mvScan.Actions->Datacap.Libraries.mvScan.Actions->scan"
141 05:35:04.399 (0) Cache file: C:\Users\Administrator\AppData\Local\Temp\DCmvscan374895062.xml
142 05:35:04.399 (0) Demo mode enabled
143 05:35:04.399 (0) Refresh cache
144 05:35:04.399 (0) Scan folder: C:\Datacap\ClaimForm\images
145 05:35:04.399 (0) Scan folder: C:\Datacap\ClaimForm\images
146 05:35:04.400 (0) Found 2 files
147 05:35:04.400 (0) Check file 1-ClaimForm-BC.tif modified 4/1/2020 6:30:15 PM
148 05:35:04.400 (0) Add file to list: 1-ClaimForm-BC.tif input folder: C:\Datacap\ClaimForm\images
149 05:35:04.400 (0) Check file 2-ClaimForm-BC.tif modified 4/1/2020 6:30:15 PM
150 05:35:04.401 (0) Add file to list: 2-ClaimForm-BC.tif input folder: C:\Datacap\ClaimForm\images
151 05:35:04.401 (0) Found 0 folders
152 05:35:04.402 (0) Wrote cached file list to C:\Users\Administrator\AppData\Local\Temp\DCmvscan374895062.xml
153 05:35:04.402 (0) Root folders contain 2 selected files and folders
154 05:35:04.402 (0) Try to ingest 2 files and folders
155 05:35:04.402 (0) Start ingest file 0 of 2
156 05:35:04.402 (0) Ingest file C:\Datacap\ClaimForm\images\1-ClaimForm-BC.tif
157 05:35:04.402 (0) Ingesting C:\Datacap\ClaimForm\images\1-ClaimForm-BC.tif
158 05:35:04.402 (0) Demo mode enabled
159 05:35:04.402 (0) Ingest to C:\Datacap\ClaimForm\batches\20200408.000007\TM000001.tif
160 05:35:04.402 (0) Demo mode enabled
161 05:35:04.409 (0) Ingested page 1
162 05:35:04.410 (0) Deleted lock file C:\Datacap\ClaimForm\images\1-ClaimForm-BC.tif.lock
163 05:35:04.410 (16) Start ingest file 1 of 2
164 05:35:04.411 (0) Ingest file C:\Datacap\ClaimForm\images\2-ClaimForm-BC.tif
165 05:35:04.411 (0) Ingesting C:\Datacap\ClaimForm\images\2-ClaimForm-BC.tif
166 05:35:04.411 (0) Demo mode enabled
167 05:35:04.412 (0) Ingest to C:\Datacap\ClaimForm\batches\20200408.000007\TM000002.tif
168 05:35:04.412 (0) Demo mode enabled
169 05:35:04.413 (0) Ingested page 2

```

Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-6. Sample log file - task was not completed successfully

The screen capture shows a sample vscan_rrs.log file in Notepad++. The text indicates that the task was not completed successfully.

The scan() action is defined in the VScan ruleset and is run for the VScan task.

Observe the following information on the log file:

- The set_folder action is run successfully as indicated by the text: action returned true and Image source folder set to: C:\Datacap\ClaimForm\images
- The set_types and set_copy_folder actions are run successfully as indicated by the text: File types to be ingested: tif and Copy folder set to: C:\Datacap\ClaimForm\images
- The scan action is not run successfully as indicated by the text: Found 0 files, Found 0 folders
- The text Ingested page 1 and Ingested page 1 that you saw in the previous log file when the batch was run successfully is missing in this file.
- The message indicates that no files were found in the image folder and no files were ingested.

6.2. Troubleshoot the applications in the Datacap Studio Test tab

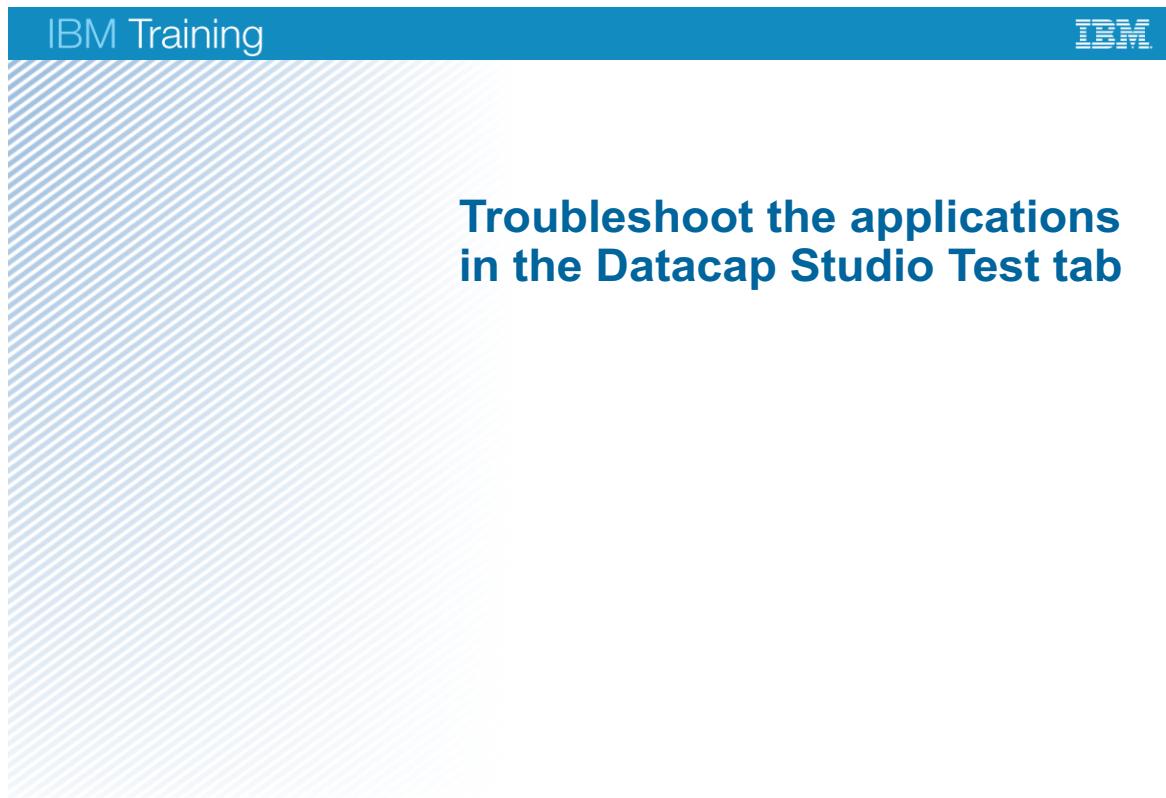
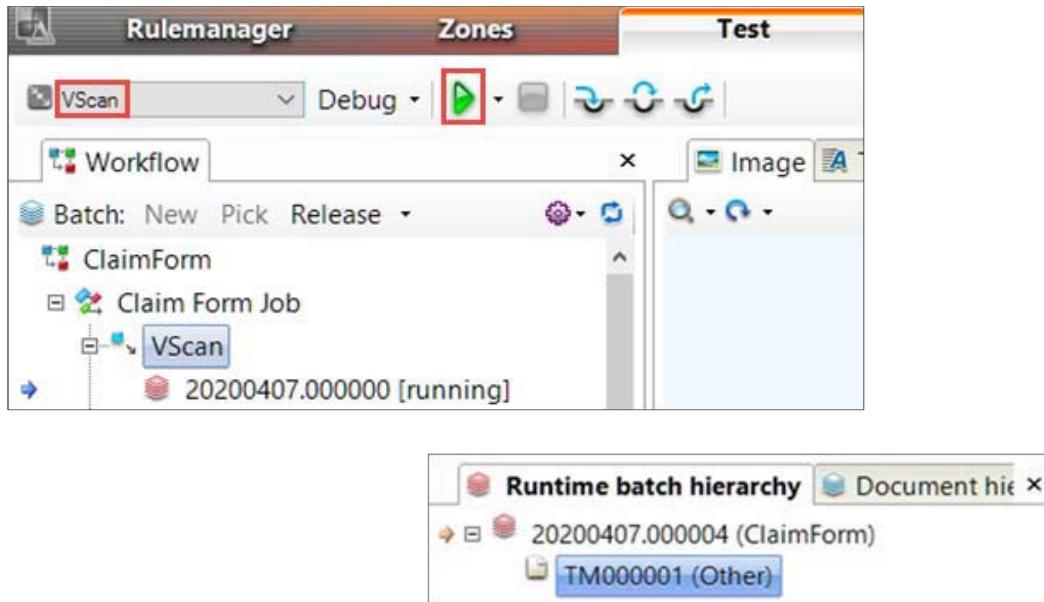


Figure 6-7. Troubleshoot the applications in the Datacap Studio Test tab

IBM Training 

Testing the application at design time



Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-8. Testing the application at design time

In the previous unit, you learned about how to test your application at design time within Datacap Studio.

The screen captures show the Test view in Datacap Studio.

- The upper screen capture shows that a batch is running at the first task (VScan) in the Workflow pane. A batch item is shown under the VScan node.
- The lower screen capture shows that the scanned page is listed in the Runtime batch hierarchy. At this step, the page type is shown as Other.

You can run all the tasks in a workflow in Datacap Studio Test tab.

Datacap Studio Test tab debugging features

- Using Breakpoints
 - Halt batch process when Rulerunner encounters the specified element: ruleset, rule, or action, document, page, or field
- Single-stepping through your code
 - Determine whether the functions and actions work
 - Actions return True (check mark) and False (exclamation point)
- Examining log files from the Test tab

Figure 6-9. Datacap Studio Test tab debugging features

Using Breakpoints

A breakpoint stops batch process when Rulerunner encounters the specified element.

- A breakpoint stops at a predetermined ruleset, rule, or action.
- A breakpoint can also stop the task when a task starts processing a specific document, page, or field.

The document, page, or field must exist in the runtime hierarchy before you can set a breakpoint on it.

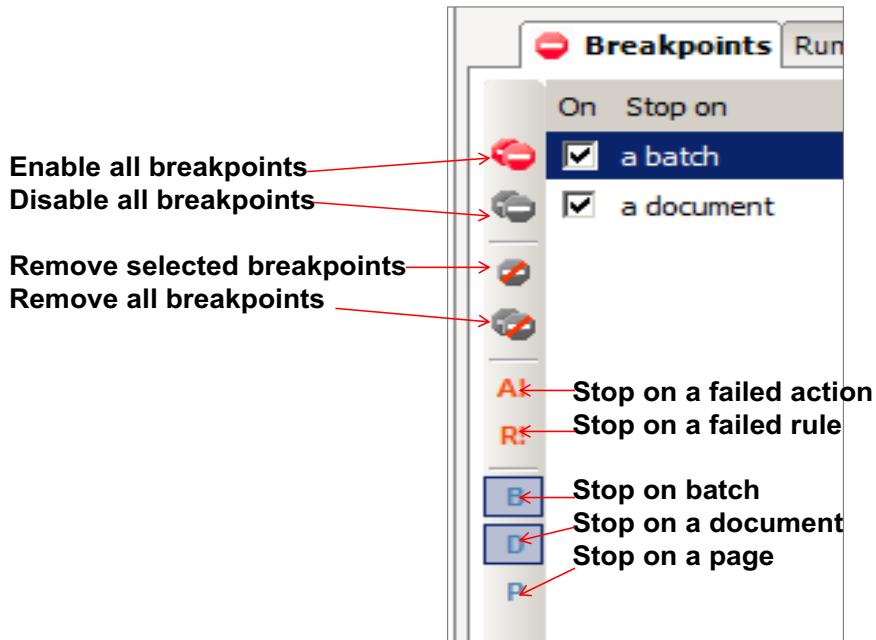
Single-stepping

- It is useful to determine whether the functions and actions within a rule are operating as intended
- As you step through each line, you can see the actions that returned True (check mark) and False (exclamation point)

Examining log files from the Test tab

You can access the Output tab in Datacap Studio to view the output that is written to different log files.

Breakpoint pane in the Datacap Studio Test tab



Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-10. Breakpoint pane in the Datacap Studio Test tab

The screen capture shows the Breakpoints pane in the Datacap Studio Test tab. The pane contains the tools that are available for setting breakpoints.

Setting generic breakpoints

- The Datacap Test tab includes two controls that you can select to halt processing when any rule or action fails.

Control	Command	Description
A!	Stop on a failed action	Click this button to add a generic breakpoint that halts processing whenever an action fails
R!	Stop on a failed rule	Click this button to add a generic breakpoint that halts processing whenever a rule fails

Figure 6-11. Setting generic breakpoints

The table on the slide illustrates the generic breakpoint run options:

A! – Stop on a failed action

R! – Stop on a failed rule



Setting specific breakpoints on batch hierarchy

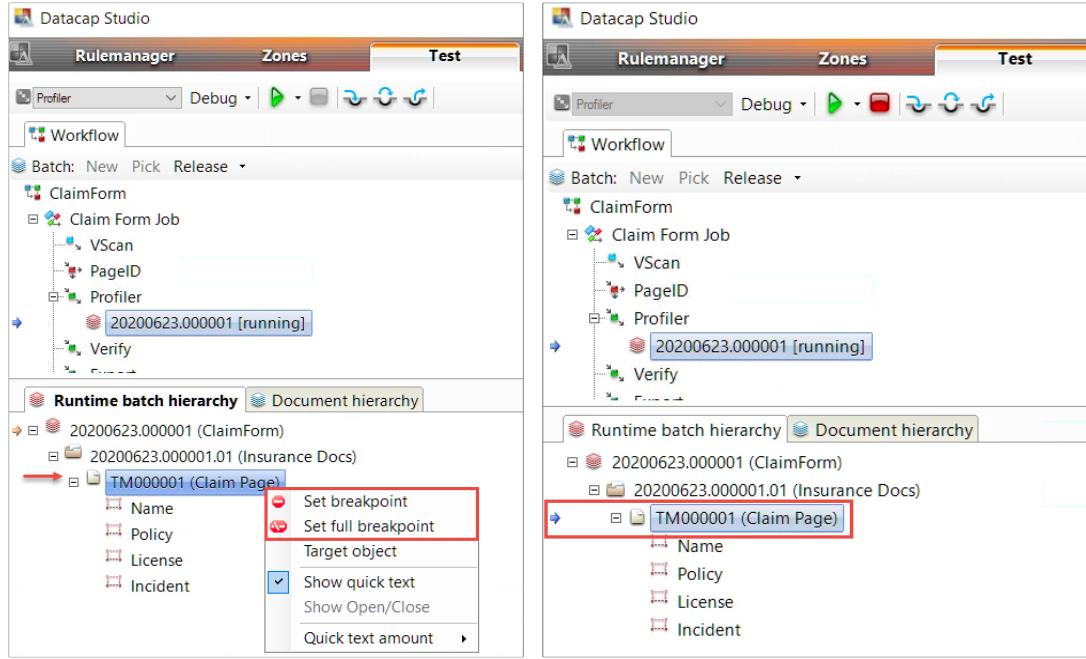


Figure 6-12. Setting specific breakpoints on batch hierarchy

The screen captures show the Datacap Studio Test tab. The left screen capture shows the menu for a page object (Claim Page) in the Runtime batch hierarchy pane. The Set breakpoint menu items are highlighted.

The right screen capture shows the Runtime batch hierarchy pane after you set the breakpoint and run the Profiler task. The batch process stops at the Page object (Claim Page) because of the breakpoint.

To set a breakpoint on a document, page, or field:

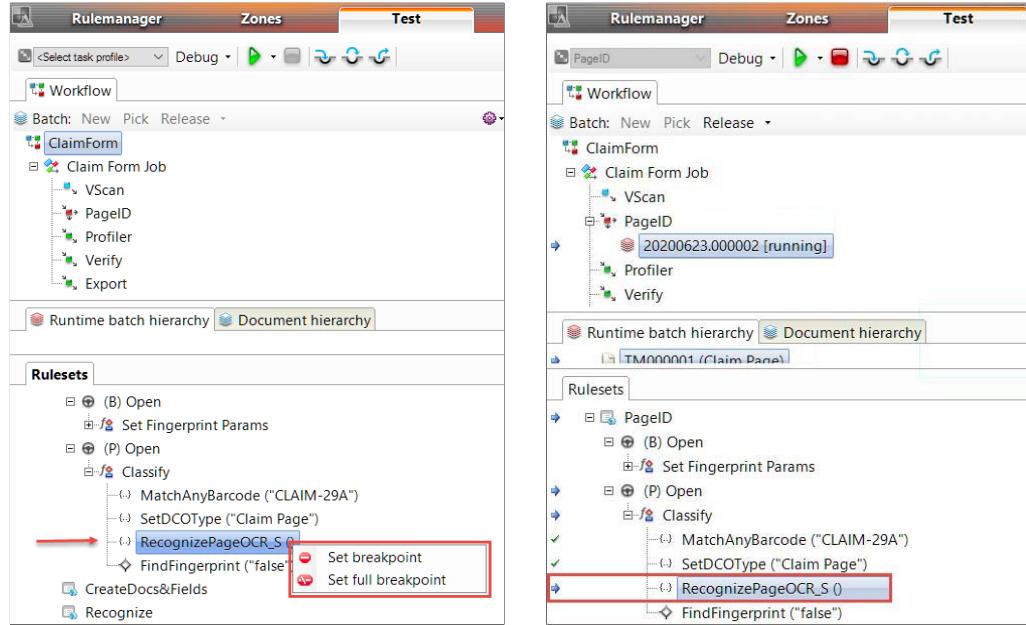
- On the Datacap Studio Test tab, select the Runtime batch hierarchy pane.
- Right-click the item and select Set breakpoint or Set full breakpoint.

There are two types of breakpoints, both of which halt execution when Rulerunner encounters the specified element.

- Breakpoints: Halts execution when the Rulerunner execution manager encounters the specified element, regardless of context.
- Full breakpoints: Halts execution when the Rulerunner execution manager encounters the specified element within the same context.



Setting specific breakpoints on rulesets, rules, or actions



Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-13. Setting specific breakpoints on rulesets, rules, or actions

The screen captures show the Datacap Studio Test tab. The left screen capture shows the menu for an action (RecognizePageOCR_S) in the Rulesets pane. The Set breakpoint menu items are highlighted.

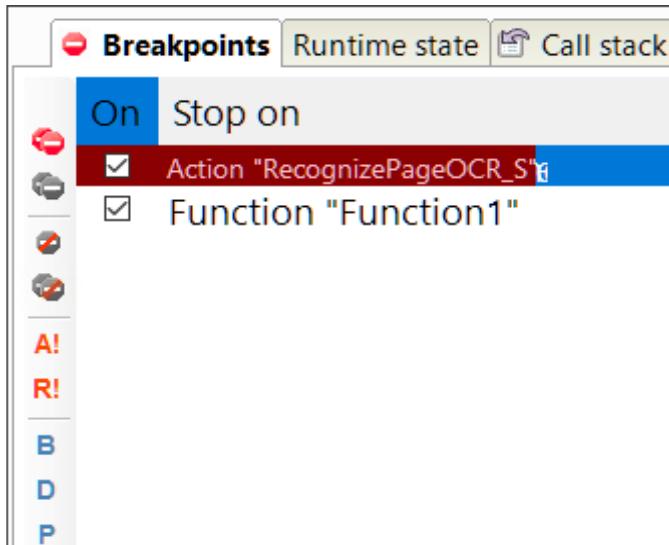
The right screen capture shows the Rulesets pane after you set the breakpoint and run the PagelD task. The batch process stops at the RecognizePageOCR_S action because of the breakpoint.

To set a breakpoint on a ruleset, rule, function, or action:

- On the Datacap Studio Test tab, select the Rulesets pane.
 - Right-click the item and select Set breakpoint or Set full breakpoint.
- For rulesets and rules, you can set only breakpoints, not full breakpoints.

Disabling and clearing breakpoints

- The Breakpoints pane displays all of the defined breakpoints
- The checkbox indicates whether the breakpoint is enabled or unavailable



Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-14. Disabling and clearing breakpoints

The screen capture shows the Breakpoints pane in Datacap Studio Test tab. A list of breakpoints that are set for the application are listed.

You can enable or disable individual breakpoints by selecting or clearing checkboxes.

The buttons on the left of the Breakpoints pane are options to enable, disable, or remove breakpoints.

The checkbox to the left of each breakpoint indicates whether the breakpoint is enabled or unavailable.

By default, breakpoints are enabled when you add them.

UI controls for Stepping through the code

- The Step in control steps into the next line of code
- The Step/Step over starts the next line of code and any lower-level functions and actions, and then stops
- The Step out control steps through the next line of code



Figure 6-15. UI controls for Stepping through the code

The screen capture shows the UI controls for stepping through code in Datacap Studio Test tab toolbar. The controls are enabled with tooltips.

Single stepping is useful to determine whether functions and actions within a rule are operating as intended.

Step in

The Step in control steps into the next line of code. If the next line calls a rule or function, Step in opens the rule or function and halts inside it. If the next line is an action, Step in opens the action. You must click it again to close the action.

Step/Step over

It starts the next line of code and any lower-level functions and actions, and then stops. If the next line is an action, Step over works like Step in and opens the action.

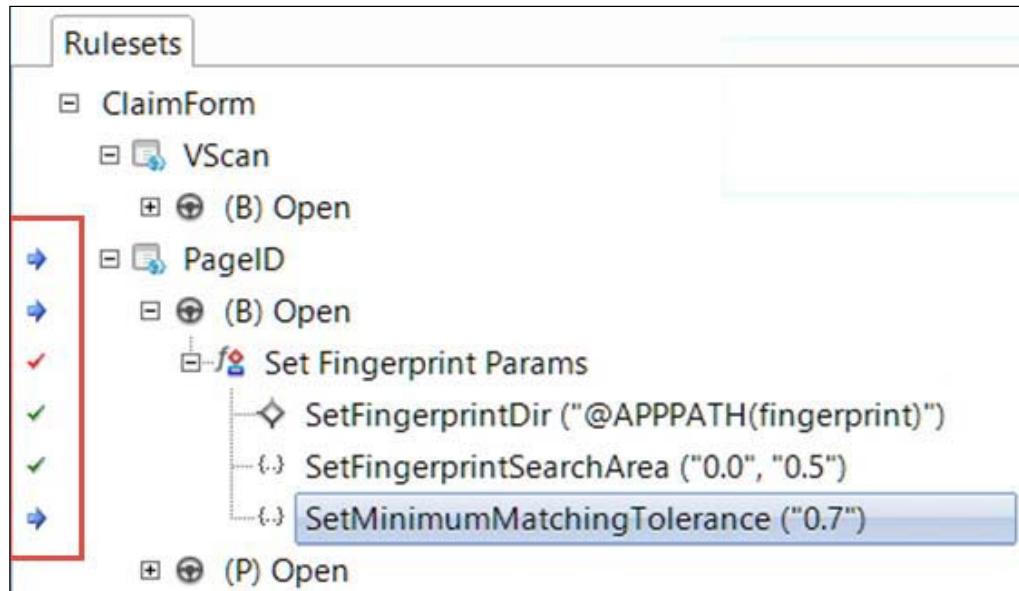
Step out

It steps through the next line of code. If the next line is a rule or function, Step out works like Step over and starts any lower-level functions and actions.

If the next line is an action, Step out starts and closes the action.

IBM Training IBM

Single stepping through a batch



Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-16. Single stepping through a batch

The screen capture shows the Rulesets pane in Datacap Studio Test tab. Single step in for a batch is shown.

The Step in control steps into the next line of code. As you step through each line, you can see the actions that are returned as True (check mark) or False (exclamation point). If an action returns false, you can look in the batch log to see why the action returned false.

You can also access the Output tab in Datacap Studio to view the output that is written to different log files



Examining log files from the Test tab

The screenshot shows the Datacap Studio interface with the 'Test' tab selected. The left sidebar contains 'Workflow' and 'Profiler' sections. The 'Runtime batch hierarchy' pane shows a tree structure for '20200623.000002 (ClaimForm)'. The 'Rulesets' pane shows 'ClaimForm' and 'VScan'. The 'Output' tab is selected in the top navigation bar. A red box highlights the 'Batch log' section, which displays a list of log entries. The log entries show timestamp, process ID, action, and details. For example, it shows rule starts, function executions, and action successes.

```

22:14:13.541 (0) t:2868 p:19ACE50 /execute statement On Rule Start
22:14:13.541 (0) t:2868 p:19ACE50 func "Create Fields"
22:14:13.541 (0) t:2868 p:19ACE50 execute statement On Function St
22:14:13.541 (0) t:2868 p:19ACE50 executing code:
22:14:13.541 (0) t:2868 p:19ACE50 g_Ftmr=cdbl(Timer)
22:14:13.541 (0) t:2868 p:19ACE50 /execute statement On Function S
22:14:13.541 (0) t:2868 p:19ACE50 calling action CreateFields () on P
22:14:13.541 (0) t:2868 p:19ACE50 execute statement On Action St
22:14:13.541 (0) t:2868 p:19ACE50 executing code:
22:14:13.541 (0) t:2868 p:19ACE50 Call OnActionStart()
22:14:13.542 (0) t:2868 p:19ACE50 Current Locale: '1033 (en-US)' English (U
22:14:13.542 (0) t:2868 p:19ACE50 /execute statement On Action S
22:14:13.542 (0) t:2868 p:19ACE50 call "Datacap.Libraries.Applicati
22:14:13.544 (0) t:AB4 p:19ACE50 Creating fields for ID: 'TM000001' Level: 2
22:14:13.544 (0) t:AB4 p:19ACE50 Success Num of Children: 4
22:14:13.544 (0) t:2868 p:19ACE50 result 0[x0] = true
22:14:13.544 (0) t:2868 p:19ACE50 /call
22:14:13.544 (0) t:2868 p:19ACE50 action returned true
22:14:13.544 (0) t:2868 p:19ACE50 execute statement On Action Tr
22:14:13.544 (0) t:2868 p:19ACE50 executing code:
22:14:13.544 (0) t:2868 p:19ACE50 Call OnActionSuccess()
22:14:13.545 (0) t:2868 p:19ACE50 Action esec='0.00000'
22:14:13.545 (0) t:2868 p:19ACE50 /execute statement On Action Ti
22:14:13.545 (0) t:2868 p:19ACE50 /action
22:14:13.545 (0) t:2868 p:19ACE50 func result: "true"

```

© Copyright IBM Corporation 2020

Figure 6-17. Examining log files from the Test tab

The screen capture shows the Output tab in Datacap Studio Test tab. You can access the information that is written to different log files on this tab. In this example, the Batch Log is selected from the list of available logs.

The output pane refreshes automatically when:

- You stop at a breakpoint
- You single-step through a line of code
- When the current task profile completes

You need to scroll to the bottom each time to see the latest messages.

Unit summary

- Examine the logs files for Datacap applications
- Learn how to troubleshoot the applications in the Datacap Studio

Figure 6-18. Unit summary

Review questions

1. True or False: Rulerunner writes log information to a Rulerunner Service log file each time it runs an action.
2. In the Datacap Studio Test tab, single-stepping is useful for which of the following scenario?
 - A. To determine whether the functions and actions within a rule are operating as intended
 - B. To stop the batch at a predetermined ruleset, rule, or action
 - C. To view output that is written to the log files
 - D. To identify the current task profile
3. True or False: In the Datacap Studio Test tab, you can view the information that is written to log files.



Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-19. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. True or False: Rulerunner writes log information to a Rulerunner Service log file each time it runs an action.
The answer is True.

2. In the Datacap Studio Test tab, single-stepping is useful for which of the following scenario?
 - A. To determine whether the functions and actions within a rule are operating as intended
 - B. To stop the batch at a predetermined ruleset, rule, or action
 - C. To view output that is written to the log files
 - D. To identify the current task profile**The answer is A.**

3. True or False: In the Datacap Studio Test tab, you can view the information that is written to log files.
The answer is True.



Figure 6-20. Review answers

Exercise: Troubleshooting a Datacap application

Troubleshooting a Datacap application

© Copyright IBM Corporation 2020

Figure 6-21. Exercise: Troubleshooting a Datacap application

Exercise objectives

- Examine the logs files for debugging
- Troubleshoot the applications in the Datacap Studio Test tab



Figure 6-22. Exercise objectives

Unit 7. Building an application to process multiple page types

Estimated time

00:40

Overview

This unit describes how to build a Datacap application for processing multiple documents of different page types in a single batch.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Describe how to create an application by using an existing application
- Learn how to configure the application in IBM Content Navigator
- Identify the document hierarchy objects
- Learn how to configure rulesets
- Describe how to configure fingerprinting and zones

Figure 7-1. Unit objectives

Topics

- Copy an existing application
- Configure the application in Datacap Navigator
- Define document hierarchy and configure rulesets
- Configure fingerprinting and zones
- Test the application

Figure 7-2. Topics

7.1. Copy an existing application

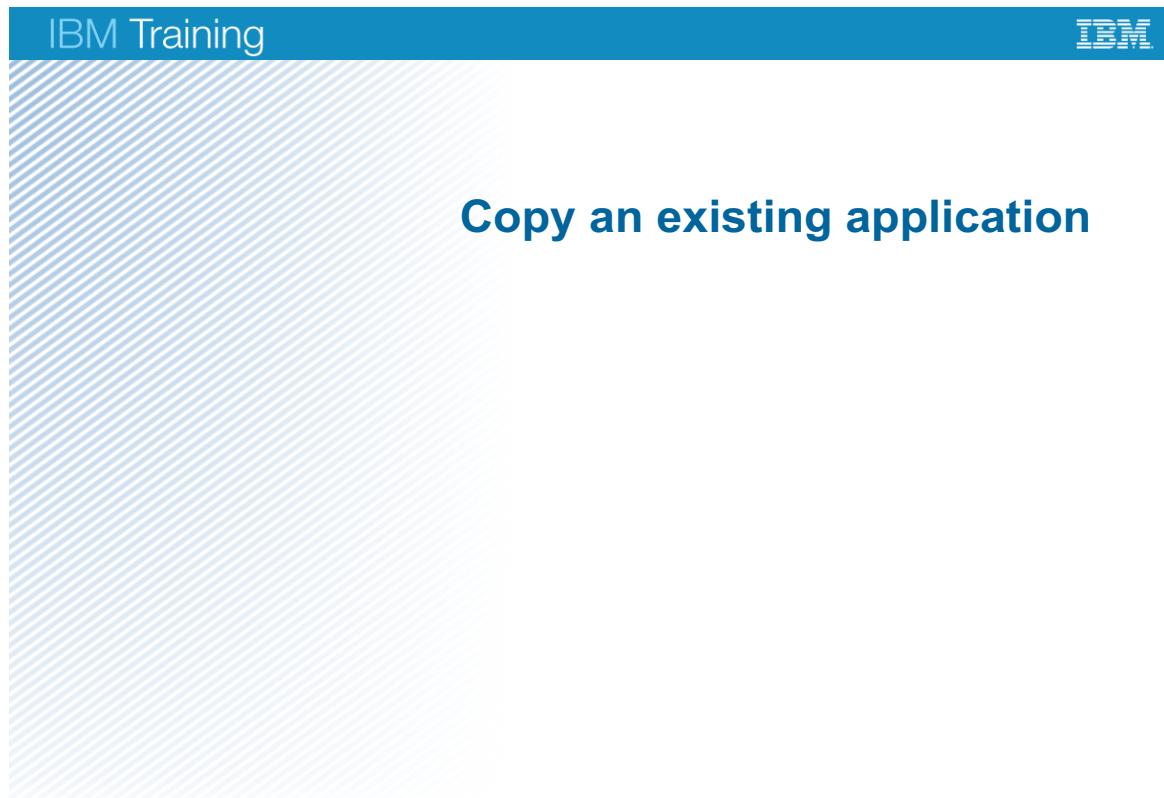


Figure 7-3. Copy an existing application

Scenario for the application

- You want to enhance a Datacap application to process a second page: Car Rental > Rental Form
- Extract values for the following fields:
 - Pickup date
 - Return date
 - Total
 - Options
- Document classify type
 - Fingerprint
- Fields extraction type
 - Fingerprint & zoning

Car Rental #1		Rental Agreement	
Pickup	Return	Car Type	Options
Mon, Oct 4, 2010 10:00AM New York (JFK)	Fri, Oct 8, 2010 5:00PM New York (JFK)	Full size	GPS Navigation <input checked="" type="checkbox"/> Child Seat <input type="checkbox"/> Fuel Service <input checked="" type="checkbox"/>
Daily Rate:	\$65.56	Optional Insurance:	\$104.95
Taxes and Fees:	\$150.02	Total:	\$582.77

Figure 7-4. Scenario for the application

You already created a Datacap application to extract data from a single page. You want to enhance the application to process multiple pages.

The documents for scanning are on the file system folder.

Document type: Car Rental

Page type: Rental Form

In this example, you extract values for the following fields:

- Pickup date
- Return date
- Total
- Options

In this example, assume that the field positions on the page are constant for all the forms of this type.

Since the fields positions are constant for this image, you can use fingerprints and zones to identify the page and extract the value from the fields.

Steps to configure the Datacap application

Steps	Datacap tools
• Copy an application	Datacap Studio
• Configure jobs and workflow	Datacap Navigator
• Complete application settings	Application Manager
• Define DCO, create rulesets, configure task profiles, configure fingerprints and zones for the application	Datacap Studio
• Create Rulerunner threads	Rulerunner Manager

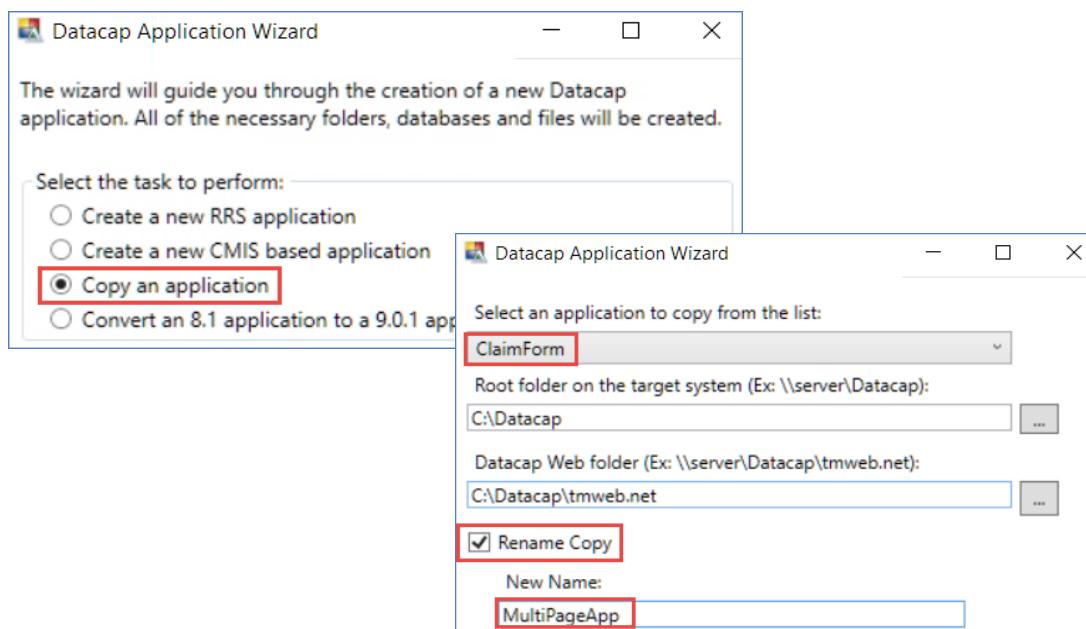
Figure 7-5. Steps to configure the Datacap application

Steps to build an application

1. In Datacap Studio, create a copy of an existing application that is based on the Forms template.
2. In Datacap Navigator admin tool, configure workflow and jobs.
3. In Datacap Application Manager, complete the application settings
4. In Datacap Studio, configure the application: Define Datacap hierarchy (DCO), create rulesets, configure task profiles, configure fingerprinting & zones, and test the application
5. In Datacap Rulerunner Manager, create rulerunner threads to run the background tasks.

IBM Training 

Copy an existing application to create a new application



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-6. Copy an existing application to create a new application

The upper screen capture shows the options that are available to create a new application in Datacap Application wizard. The Copy an application option is highlighted. The lower screen capture shows the application name to copy, folder location where the application is added, the option to rename the copy, and the name for the new application. You start Datacap Application Wizard in Datacap Studio.

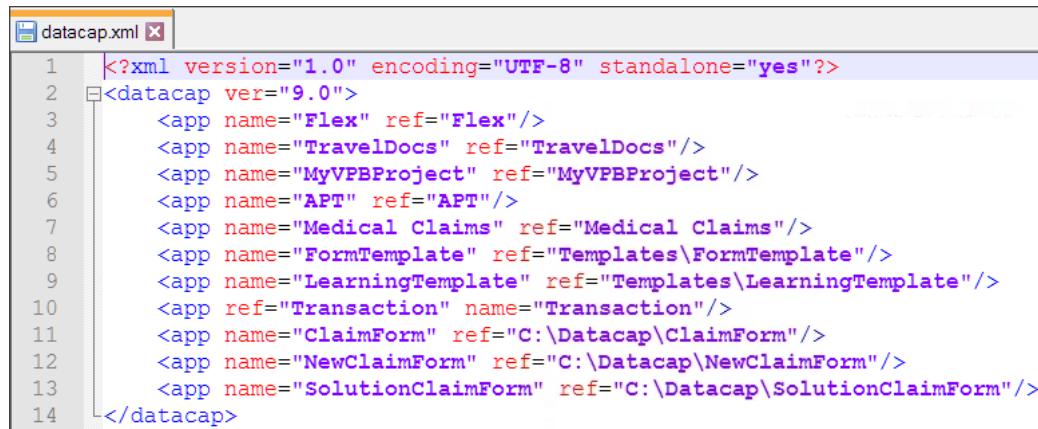
If an application exists that can be used to enhance for a new business scenario, you can copy an existing application to create a new one.

Datacap Folder

- It is the root folder for the application.
- The Datacap folder is normally on the local system but it might be on another server.
- To select a folder on another server, the path is: \\<server name>\Datacap

Datacap.xml

- Whenever a new application is created, it creates an entry in the Datacap.xml



```

1  <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2  <datacap ver="9.0">
3      <app name="Flex" ref="Flex"/>
4      <app name="TravelDocs" ref="TravelDocs"/>
5      <app name="MyVPPProject" ref="MyVPPProject"/>
6      <app name="APT" ref="APT"/>
7      <app name="Medical Claims" ref="Medical Claims"/>
8      <app name="FormTemplate" ref="Templates\FormTemplate"/>
9      <app name="LearningTemplate" ref="Templates\LearningTemplate"/>
10     <app ref="Transaction" name="Transaction"/>
11     <app name="ClaimForm" ref="C:\Datacap\ClaimForm"/>
12     <app name="NewClaimForm" ref="C:\Datacap\NewClaimForm"/>
13     <app name="SolutionClaimForm" ref="C:\Datacap\SolutionClaimForm"/>
14 </datacap>

```

Figure 7-7. Datacap.xml

The screen capture shows the Datacap.xml file in Notepad++.

The Datacap.xml file is stored in the C:\Datacap folder on the Datacap server system.

This file maintains a list of applications that are available on the system. This file is called to show the list of applications in the Datacap clients.

For every new application, Datacap adds an entry for the new application in the Datacap.xml file.

The Application wizard does not create the datacap.xml file. It adds an entry for the new application.

7.2. Configure the application in IBM Content Navigator



Figure 7-8. Configure the application in IBM Content Navigator




Create a repository in IBM Content Navigator

General Configuration Parameters

You must connect to the repository before you configure parameters, system properties, folders, and so on.

* Display name:	<input type="text" value="MultiPageApp"/>
* ID:	<input type="text" value="MultiPageApp"/>
* Datacap wTM URI:	<input type="text" value="http://ecmedu01:85/ServicewTM.svc"/>
* Application:	<input type="text" value="MultiPageApp"/> <div style="border: 2px solid red; padding: 2px; margin-top: 5px;">MultiPageApp</div>
* Default Station:	<input type="text" value="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
Connect...	

Building an application to process multiple page types

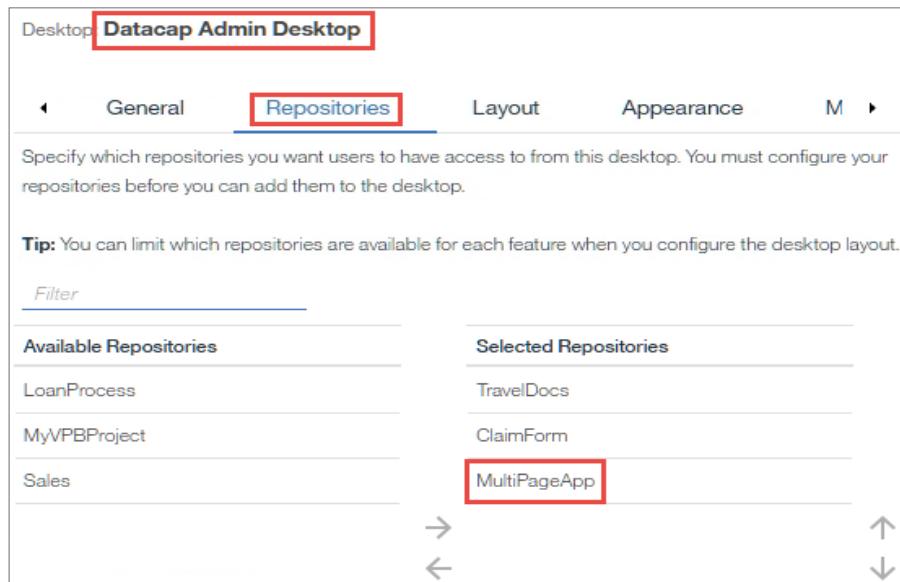
© Copyright IBM Corporation 2020

Figure 7-9. Create a repository in IBM Content Navigator

The screen capture shows the configuration of a Datacap application as a repository in the IBM Content Navigator administration tool. You provide the wTM URI to access the Datacap Server.

As you learned in a previous unit, you need to configure the Datacap application in the IBM Content Navigator administration tool. Since the Datacap Navigator clients are built on the IBM Content Navigator framework, the configuration is required to access applications in the Datacap Navigator clients.

Add a repository to the Datacap admin desktop



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-10. Add a repository to the Datacap admin desktop

The previous slide shows how to configure a Datacap application as a repository in the IBM Content Navigator administration tool. The repository must be associated with the Datacap Admin Desktop that is the Datacap Navigator admin console so that the new application is accessible in the Datacap Navigator clients.

The screen capture shows the Repositories tab for the Datacap Admin Desktop in the IBM Content Navigator administration tool.

- On this tab, you can associate a repository for a Datacap application.
- The Available Repositories pane lists all the repositories that are configured for the desktop.
- For the users to have access, the repositories must be moved to the Selected Repositories pane.

The screenshot shows the Datacap Admin Console interface. At the top, there's a blue header bar with the text "IBM Training" on the left and the IBM logo on the right. Below the header, the main title "Select your application in Datacap Administration Console" is displayed in a large, bold, blue font. A bulleted list provides instructions: "To edit your Datacap application in Datacap Administration Console, select it from the list." On the left side, there's a sidebar with links for "Datacap Navigator", "Workflows" (which is currently selected and highlighted in blue), "Groups", "Users", and "Stations". The main content area has a title "Datacap Admin Console" with a "Workflows" icon. Below the title, there's a sub-section titled "Workflows" with the sub-instruction "Configure workflow properties, jobs, and tasks." At the bottom of this section are three buttons: "Edit", "Refresh", and "Close". To the right of the main content area, there's a search dropdown menu with the placeholder text "Name contains". Inside the dropdown, several application names are listed: "MultiPageApp", "TravelDocs", "ClaimForm", "MyVPBProject", and "MultiPageApp" again. The second "MultiPageApp" entry is highlighted with a red rectangular box. At the very bottom of the screen, there are two copyright notices: "Building an application to process multiple page types" and "© Copyright IBM Corporation 2020".

Figure 7-11. Select your application in Datacap Administration Console

The screen capture shows the Datacap Admin Console where you can configure the workflow jobs and tasks for Datacap applications. To edit your Datacap application, select it from the list.

IBM Training 

Rename a job in Datacap Admin Console

Job: Claim Form Job

General	Tasks
* Name 	MultipageApp Job
Description	Standard processing of single page images
* Priority 	5 

Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-12. Rename a job in Datacap Admin Console

The screen capture shows the configuration page for a Job in the Datacap Navigator Admin Console.

When you copy an existing application, the workflow jobs and tasks from the source Datacap application might be applicable for the new application. In this case, the only change that is required is the name of the job.

7.3. Define document hierarchy and configure rulesets

Define document hierarchy and configure rulesets

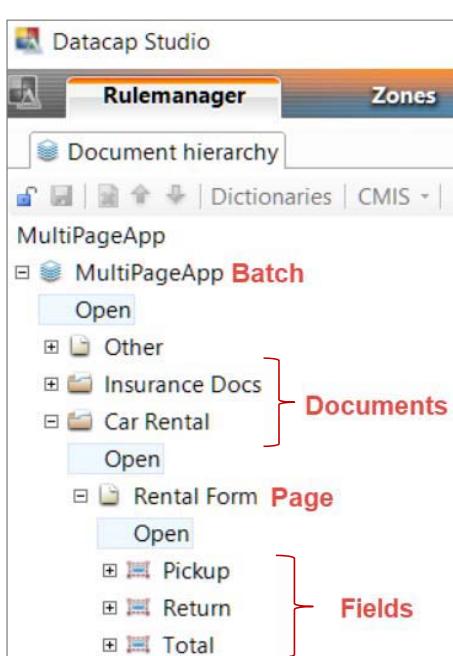
Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-13. Define document hierarchy and configure rulesets

IBM Training 

Define Document hierarchy objects



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-14. Define Document hierarchy objects

The screen capture shows the Document hierarchy (also known as DCO) pane in Datacap Studio Rulemanager tab. On this pane, you can configure the DCO (Document, Page, and Field) objects based on your business needs.

In the previous unit, you learned about the application design plan. Before you develop the application, review the documents that are used for scanning and decide the names for the Document, Page, and Field objects. The names for the application, document, page, or field objects must be unique. You receive an error when you try to enter the same name.

The component levels within the hierarchy are Batch, Document, Page, and Field.

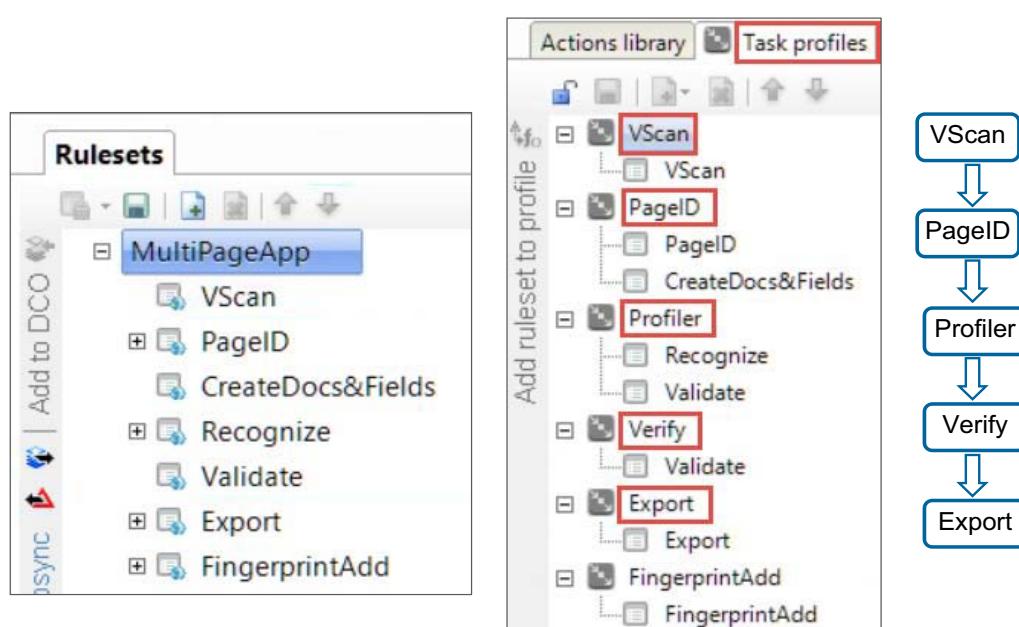
- For each application, batch (Example: MultiPageApp) is the root level object.
- Each batch contains one or more document types (Example: Insurance Docs, Car Rental)
- Each document type has at least one page type (Example: Rental Form)
- Each page type has a number of fields (Examples: Pickup, Return, and Total)

Each component has an Open and a Close node in the hierarchy. The Open and Close nodes are used to assign processing rules to the beginning or the end of each component.

After you define the Document hierarchy (DCO) and configure rules that define what actions to run, you associate the rule with a DCO. The association indicates what actions can run on which Datacap object.

IBM Training 

List of rulesets and task profiles for the application



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-15. List of rulesets and task profiles for the application

The leftmost screen capture shows an example list of rulesets in the Rulesets pane in Datacap Studio. These rulesets are used for this unit.

The rightmost diagram shows the steps for the sample workflow. Compare them with the Task profiles in the middle screen capture.

The rulesets are associated with task profiles. Compare the rulesets in the Rulesets pane with the rulesets that are on the Task profiles tab. Each task runs the rulesets in the order it is listed in the task profile.

After you configure rulesets that define what actions to run, you associate the rulesets with task profile. This association indicates what actions run when (which step in the workflow).

To keep the rulesets organized, you can order them in the Rulesets pane to match with the order in Task profiles (although it is not required).

FingerprintAdd

This ruleset is specific name. Datacap Studio Zones tab uses it. It is not tied to the tasks in the workflow, but Datacap recognizes it.

VScan ruleset

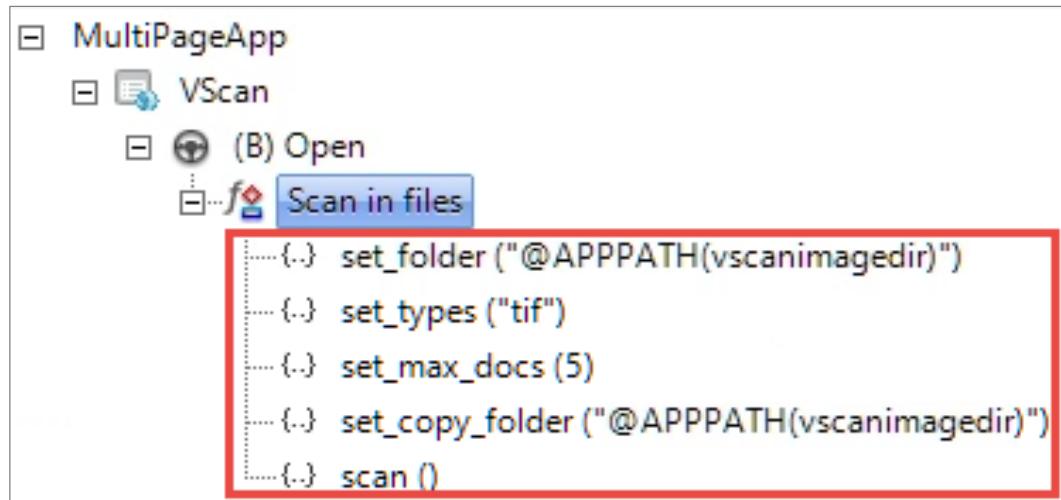


Figure 7-16. VScan ruleset

For scanning the images, you add actions to the VScan ruleset.

The example on the screen capture contains the VScan ruleset, (B) Open rule, and Scan in files function.

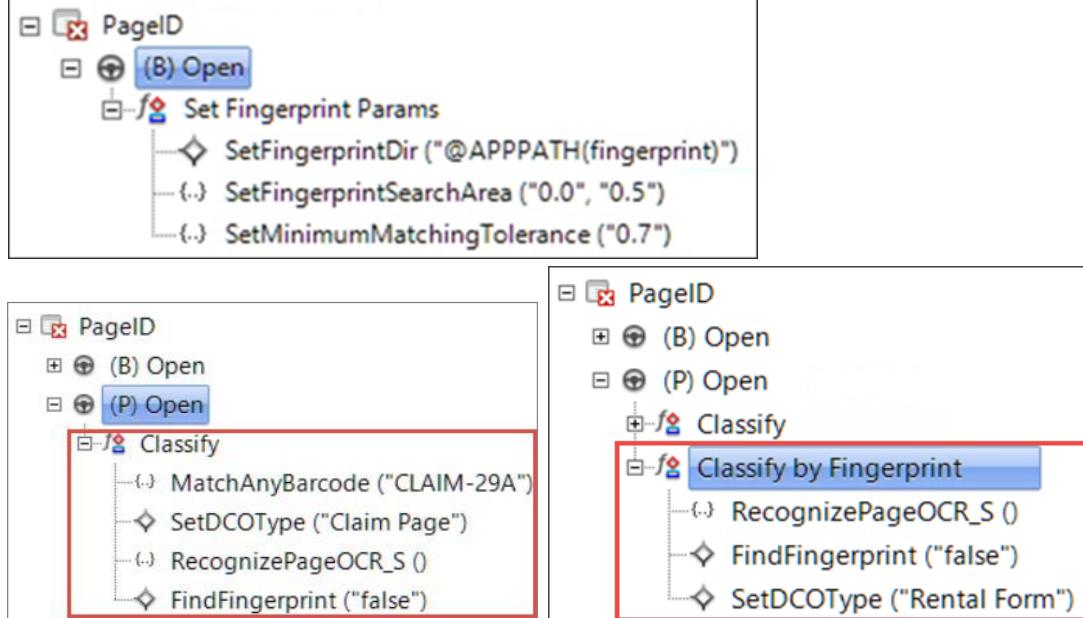
The actions in the “Scan in files” rule sets the following values:

- A folder from where images are scanned for the batch
The images folder is defined in Datacap Application Manager.
- The type of files
- Maximum number of the documents that can be scanned per batch

Because the application is copied from an existing application for the exercise, the VScan is ruleset is already configured.



Page ID ruleset



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-17. Page ID ruleset

The screen captures show the Page ID ruleset.

The upper screen capture shows the (B) Open rule and it runs at the batch level. The actions define the fingerprint settings. It is best to run the rule at batch level because it is most efficient to load the settings one time per batch. You can run it at the page level, but it is inefficient to load the settings at every page.

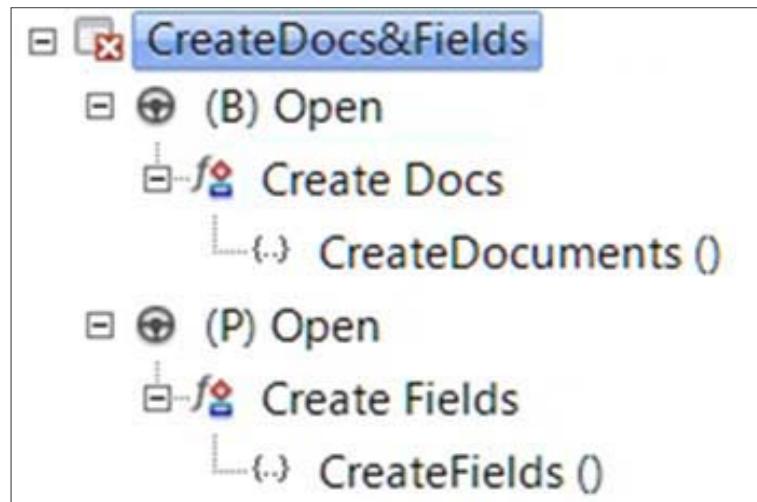
The (P) Open page-level rule contains two classify functions for recognizing two different pages:

1. Classify: shown on the lower left that uses a barcode for the page identification that contains a barcode
2. Classify by Fingerprint: shown on the lower right that uses fingerprint for the page identification that does not contain a barcode
 - The MatchAnyBarcode action searches all the barcodes on the current page and checks if any one matches the value you enter as a parameter.
 - The RecognizePageOCR_S action creates a CCO file for the current page.
 - For the FindFingerprint action, you can set true or false. If scanned image does not match the existing fingerprint, you have options to create more new fingerprints. You set false when you don't want to create more fingerprints.

- The SetDCOType action assigns a value to the Type property of the current object (Page) of the Document Hierarchy.

For the exercise, the ruleset, rules, and Classify function are defined already in the source application. After you copy the source application into a new application, you need to configure the Classify by Fingerprint function for identifying the second page type.

CreateDocs&Fields ruleset



Building an application to process multiple page types

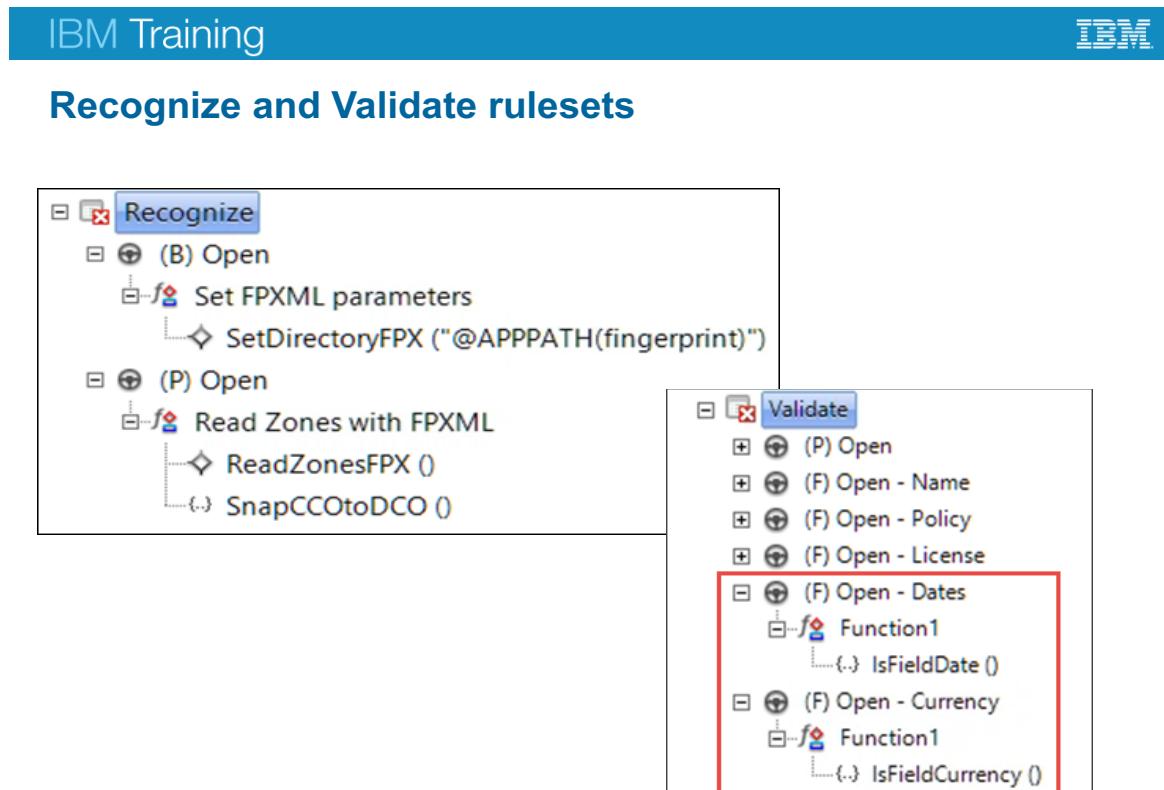
© Copyright IBM Corporation 2020

Figure 7-18. CreateDocs&Fields ruleset

The screen capture shows the CreateDocs&Fields ruleset. The actions in the CreateDocs&Fields ruleset create document, page, and fields in DCO.

The CreateDocuments action runs at the Batch level and the CreateFields action is run at the page level.

For the exercise, the ruleset comes from the source application. You need to associate the ruleset to the new page.



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-19. Recognize and Validate rulesets

The left screen capture shows the Recognize ruleset:

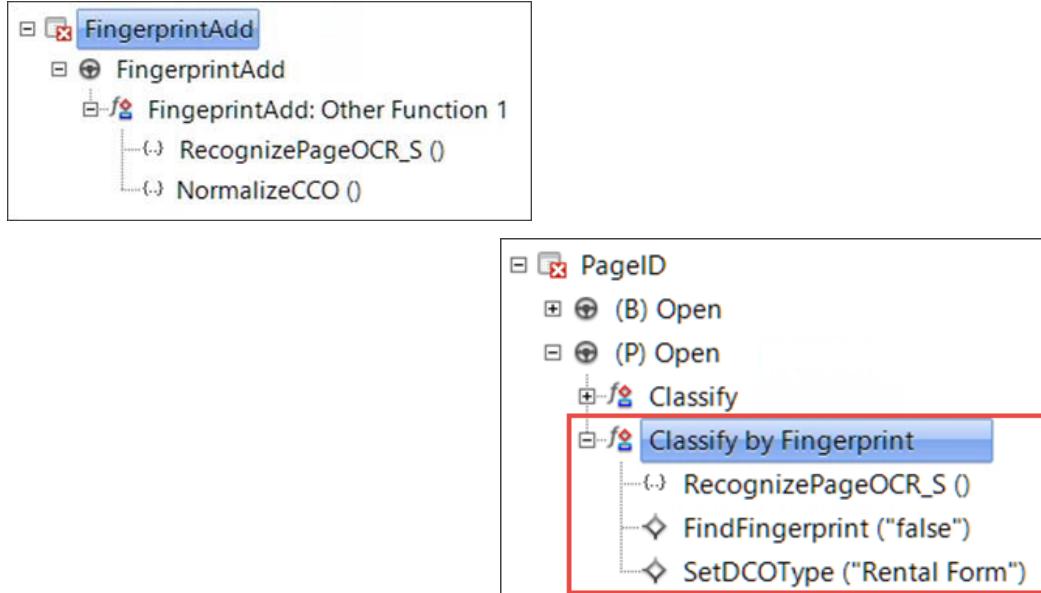
- You set the directory for FPXML at the batch level and the Read Zones action at the page level.
- The SnapCCOtoDCO() action maps the page recognition results (field values) from CCO file into DCO fields based on the zone position of the fields.

The right screen capture shows the Validate ruleset:

- The Status_Preserve_OFF action sends the verification step to human verification and it is run at the page level.
- The actions at the Field level validate the fields to check whether the field has value and the data type is correct.



Fingerprint ruleset



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-20. Fingerprint ruleset

The OCR engine that you use for the PageID ruleset to create the Fingerprint (CCO file) must match with the OCR engine that is used in the FingerprintAdd ruleset.

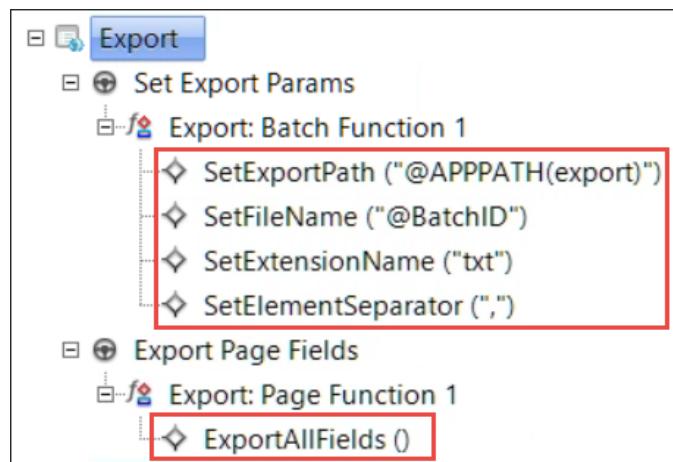
The screen captures show the FingerprintAdd ruleset on the upper left and the PageID ruleset on the lower right. Notice that both rulesets contain the same OCR action that creates the Fingerprint file: `RecognizePageOCR_S()` action from the `ocr_sr` action library.

This ruleset is added to the task profile tab, but not tied to the task in the workflow. Datacap recognizes this ruleset. In the Zones tab, each time you add a fingerprint, Datacap starts the FingerprintAdd rule set.



Export the data

- Determine where the final capture data is placed and what form it is in when it is exported
- Export data to a text file, an XML file, and other business processes



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-21. Export the data

The screen capture shows the Export ruleset for exporting the data to a text file. This ruleset is automatically added by the wizard when you create an application.

You learned about this ruleset in a previous unit.

- The actions in the first (batch level) rule of the Export ruleset sets the following parameters:
 - Folder to export file
 - File name
 - File extension
 - A separator such as comma
- The actions in the second (Page level) rule of the Export ruleset exports all the fields.

Export to a text file

The Export step is the last step of the capture process. When Capture processing is complete, you can set up Datacap to export data to a text file, an XML file, a database, a document management system, or a custom business process.

The default output format is a text file and the default export location is the exports folder for the application. The default export folder is defined in the Datacap Application Manager by the Export folder parameter.

7.4. Configure fingerprints and zones



Figure 7-22. Configure fingerprints and zones

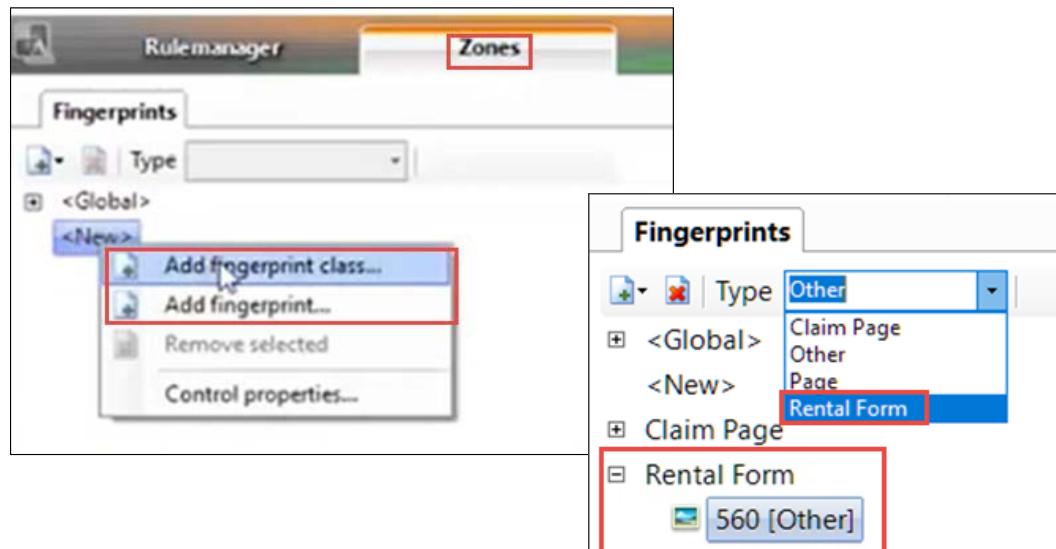
Fingerprints

- In Datacap applications, you can use fingerprints for two basic functions:
 1. Page identification to determine whether the incoming page matches a known page
 2. Identifying the field positions for each variant of each known page type
- You can add fingerprints to the fingerprint library from the Datacap Studio Zones tab.
 - Each time that you add a fingerprint, Datacap starts the FingerprintAdd rule set.

Figure 7-23. Fingerprints

Configure fingerprints in the Zones tab

1. Add a Fingerprint class
2. Add a Fingerprint - select an image and assign a page type



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-24. Configure fingerprints in the Zones tab

The screen capture shows the Zones tab. In the upper right screen capture, the Add fingerprint class and Add fingerprint actions are highlighted. The lower left screen capture shows the selection of a page type for the fingerprint.

The Add fingerprint class action creates a fingerprint class. You can add fingerprints to this class by using the Add fingerprint action.

When you add a fingerprint, it prompts you select an image file. This step runs the image file, cleans the image, and creates the Fingerprint. You can assign a type to the new fingerprint (a page type that you defined in DCO).

Fingerprint classes:

- Are helpful to differentiate the forms that you receive from various sources.
- Can be used to group fingerprints based on the source of the form.

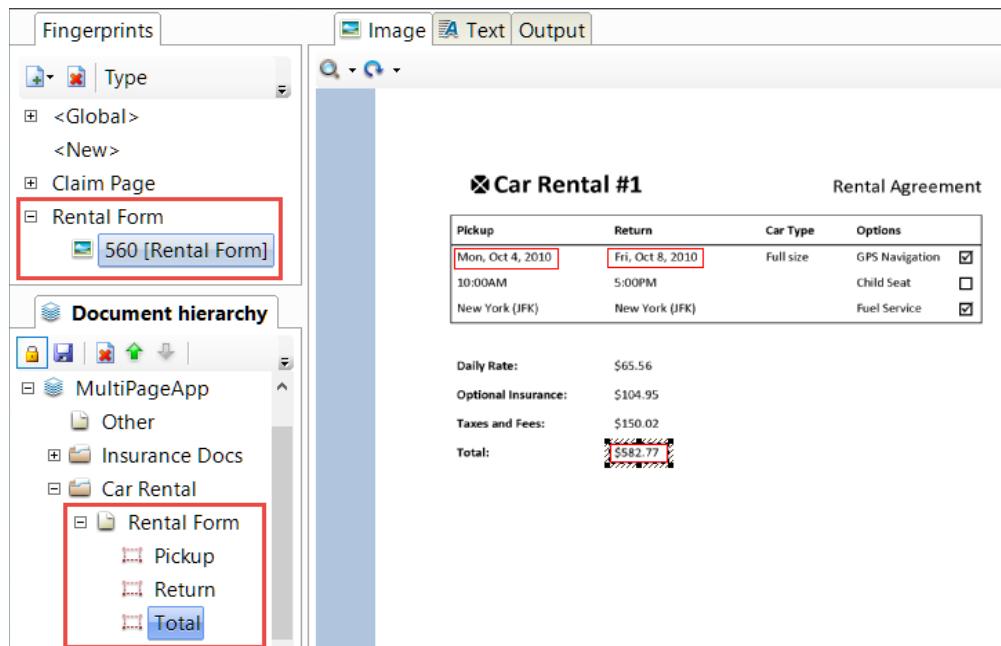
The page type names in the list come from the <application name>.xml file in the C:\Datacap\<application name>\dco_<application name>\rules folder.

Example: C:\Datacap\MultiPageApp\dco_MultiPageApp\rules\MultiPageApp.xml

The fingerprints that are available for the application is listed in the Datacap/MultiPageApp/fingerprint folder.

IBM Training

Configure zones



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-25. Configure zones

The screen capture shows the Datacap Studio Zones tab. In the Fingerprints pane, a fingerprint is selected (560[Rental Form]), and in the Document Hierarchy pane, a field (Total) is selected. On the Image tab, the image that is associated with the fingerprint (560[560[Rental Form]]) is selected and zones are drawn for the selected field.

- Fingerprints are defined to enable the identification of images by the position of data items on a page.
- After an image is identified by matching it to a fingerprint, you can extract the data from the fields that the zonal information defines in the fingerprint.
- Locating data on an image with the zonal information that is stored in the fingerprint is more efficient than other methods of locating data like doing keyword searches.

7.5. Test the application

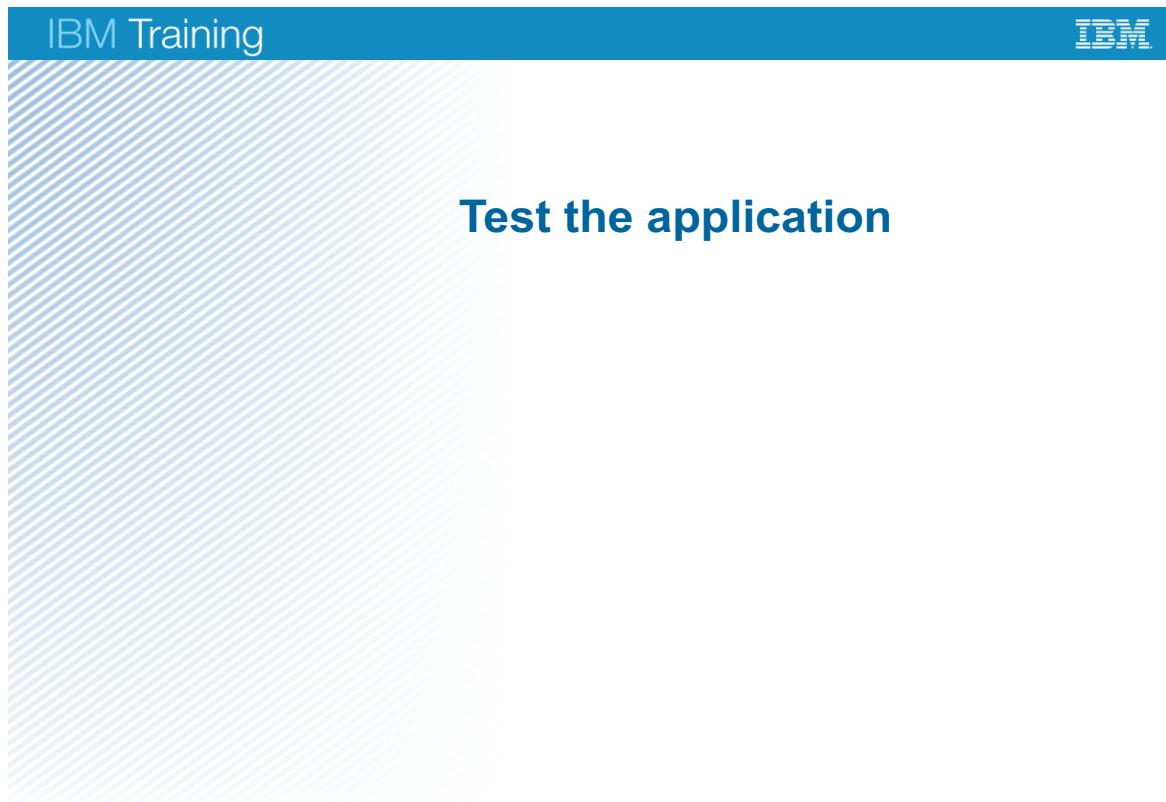


Figure 7-26. Test the application

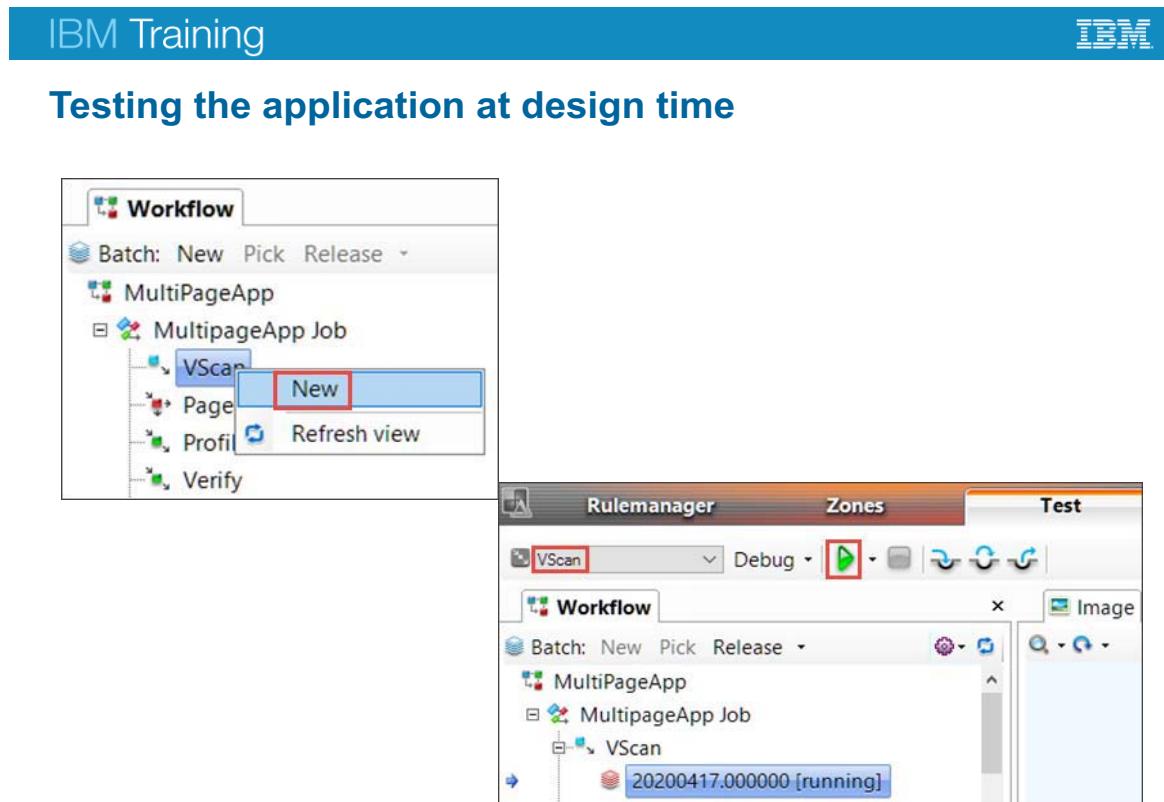
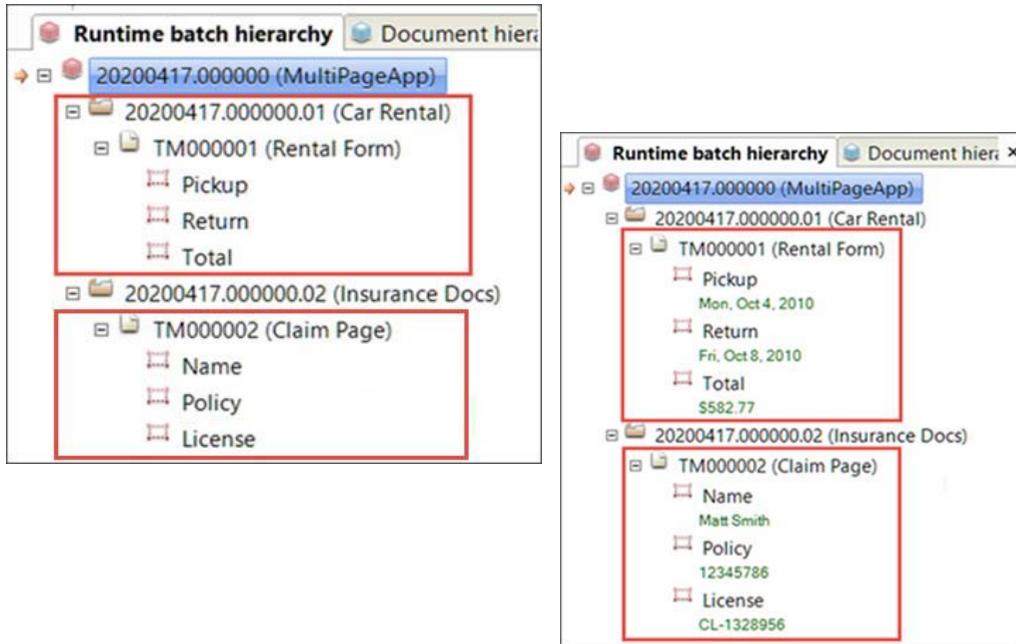


Figure 7-27. Testing the application at design time

The screen captures show the Test view in Datacap Studio.

- The upper left screen capture shows how to start a batch at the first task (VScan) in the Workflow pane.
- The lower right screen capture shows that VScan task is completed. A batch item is shown under the VScan node.

Testing the application - Page ID and Profiler tasks



Building an application to process multiple page types

© Copyright IBM Corporation 2020

Figure 7-28. Testing the application - Page ID and Profiler tasks

The screen captures show the batch results in the Runtime batch hierarchy pane of the Datacap Studio Test view.

- The left screen capture shows the results for the PageID task. The Pages (Rental Form and Claim Page) and Documents (Car Rental and Insurance Docs) types are identified. The fields are created (Pickup, Return, and Total for Rental Form and Name, Policy, and License for Claim Page)
- The right screen shows the results for the Profiler task. The extracted values are shown in DCO fields.

Testing the application - Export task

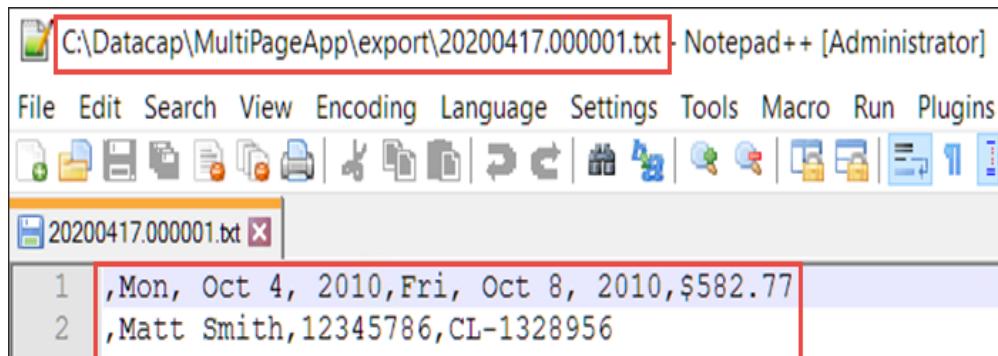


Figure 7-29. Testing the application - Export task

The screen capture shows the export text file in Notepad++. It contains two rows of comma-separated field values - one row for each page type.

The file is created after the Export task is completed for the batch.

Unit summary

- Describe how to create an application by using an existing application
- Learn how to configure the application in IBM Content Navigator
- Identify the document hierarchy objects
- Learn how to configure rulesets
- Describe how to configure fingerprinting and zones

Figure 7-30. Unit summary

Review questions

1. True or False: You can copy an existing application to create a new application by using the application wizard in Datacap Studio.
2. True or False: You cannot process documents with different types of pages in a single batch.



Figure 7-31. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. **True** or False: You can copy an existing application to create a new application by using the application wizard in Datacap Studio.
The answer is True.
2. True or **False**: You cannot process documents with different types of pages in a single batch.
The answer is False.



Figure 7-32. Review answers

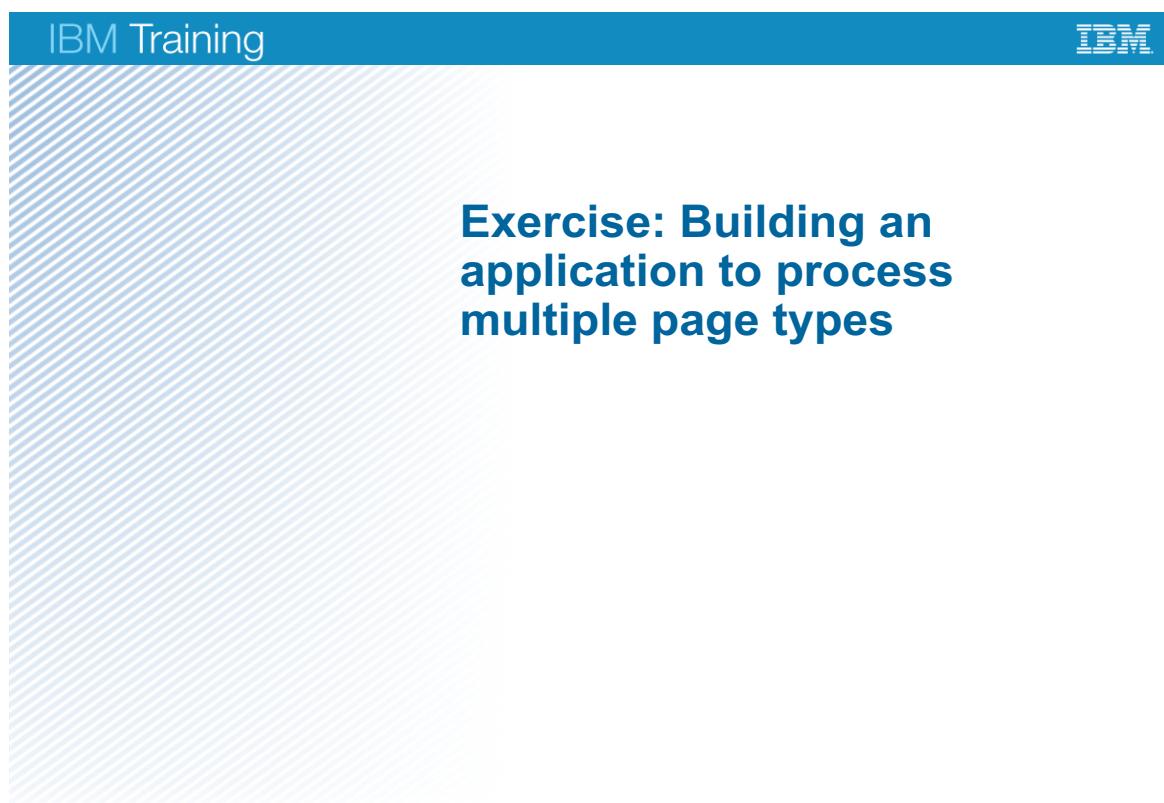


Figure 7-33. Exercise: Building an application to process multiple page types

Exercise objectives

- Create an application by using an existing application
- Configure the application in IBM Content Navigator
- Define document hierarchy and configure rulesets
- Configure fingerprinting and zones



Figure 7-34. Exercise objectives

Unit 8. Configuring an application for recognition with OMR

Estimated time

00:30

Overview

This unit teaches how to configure your application for OMR recognition and identify the batch results output.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacapoc.doc/datacap_9.1.7.htm

Unit objectives

- Learn how to configure the application for recognition with OMR
- Identify the batch results output

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-1. Unit objectives

Topics

- Configure the application for recognition with OMR
- Identify the batch results output

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-2. Topics

8.1. Configure the application for recognition with OMR



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-3. Configure the application for recognition with OMR

Optical Mark Recognition (OMR)

- Use OMR to recognize check boxes and radio buttons.
- The extracted values are shown as ones and zeros.

The figure consists of two screenshots. The left screenshot shows a form with three options: 'GPS Navigation' (checked), 'Child Seat' (unchecked), and 'Fuel Service' (checked). The right screenshot shows the Datacap interface with the runtime batch hierarchy. Under the 'Options' node in the rental form, the value '101' is displayed, indicating the state of the checkboxes.

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-4. Optical Mark Recognition (OMR)

You can use the Optical Mark Recognition (OMR) capability to recognize check boxes and radio buttons on the scan documents.

The left screen capture shows a sample multiple check boxes field with the name: Options

The right screen capture shows the extracted values from the check boxes.

After you configure the Datacap application for OMR recognition and run the batch, the data from the OMR field is exported as ones and zeros.

The ones indicate that the box is selected and zeros indicate that the box is not selected.

You can define the values to be shown as the names of the options by configuring a dictionary.

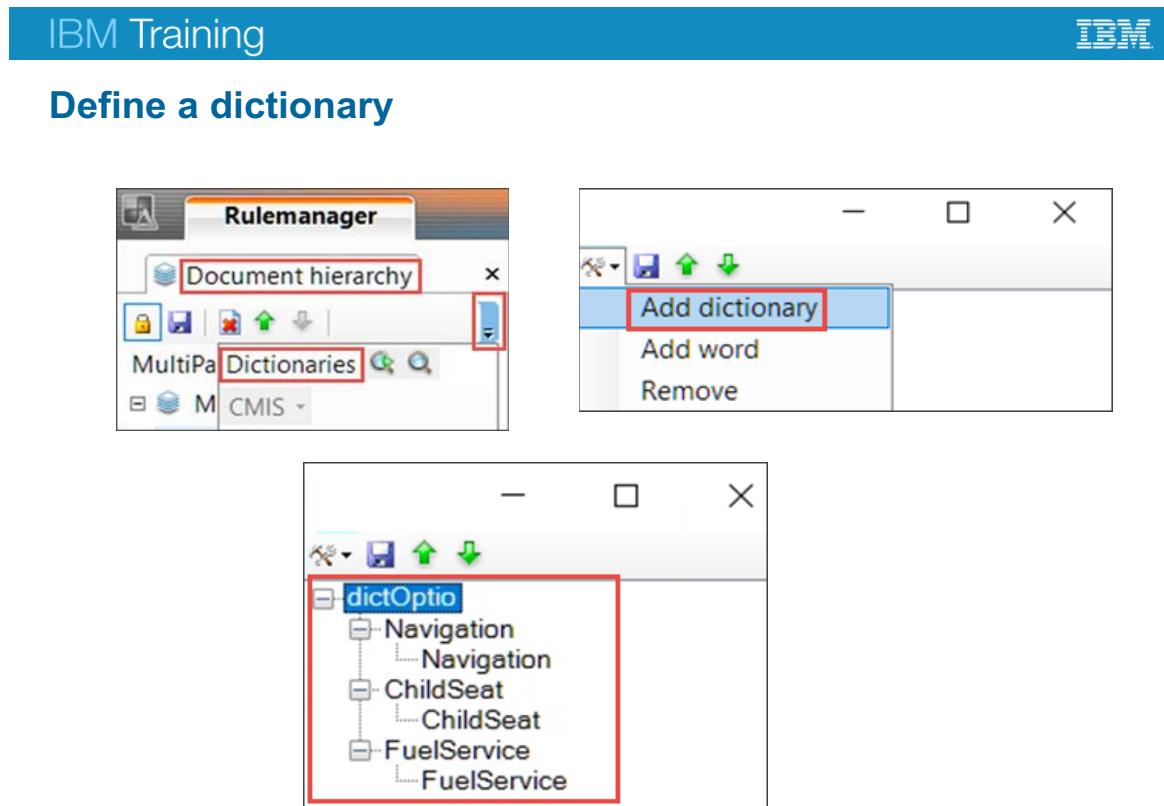


Figure 8-5. Define a dictionary

To show the extracted data as the names of the options, you define a dictionary in Datacap Studio.

The screen captures show configuration of a dictionary in Datacap Studio.

- The upper left screen capture shows Dictionaries in the Document hierarchy toolbar.
- The upper right screen capture shows the user interface for the Dictionaries tool. The menu shows the actions for adding a dictionary, a word, for deleting the objects.
- The lower screen shows a sample dictionary. The fields names are mapped with the dictionary words that you specify.



Configure variables for an OMR field

Options	
New	Remove
Object general information Max 0 Min 0 Order 0	
Variables Default_Position 1127,329,1433,570 DICT dictOptions ID 0 length MAX_TYPES 0 MIN_TYPES 0 MultiPunch 1 Pos560 1127,329,1433,570 Position 0,0,0,0 RecogType 4 ReqConf 8 rules <in /><out /> STATUS 0 TYPE Field	

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-6. Configure variables for an OMR field

The screen captures show configuration of variables for an OMR field with the name: Options

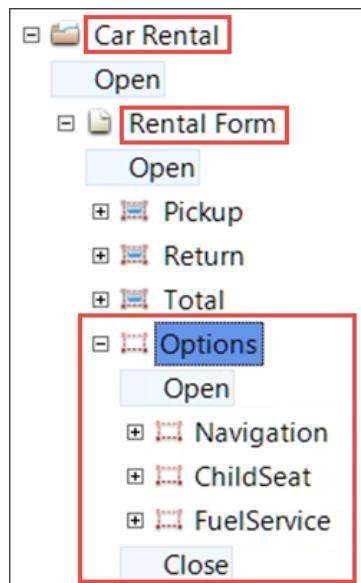
The left screen capture shows the Manage variable action in the menu in Document hierarchy.

The right screen shows the configuration page. The following variables and their values are highlighted.

- DICT: dictOptions specifies the dictionary name to use for the Options field
- MultiPunch: 1 indicates that multiple selections are allowed for the field.
- RecogType: 4 specifies the code for the recognition engine to use when reading data from this field. OMR (checkbox) fields require RecogType=4 and these fields are the only ones that typically require this variable.

You cannot use different names for the variables. Datacap recognizes the specified names.

Define document hierarchy for an OMR field



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

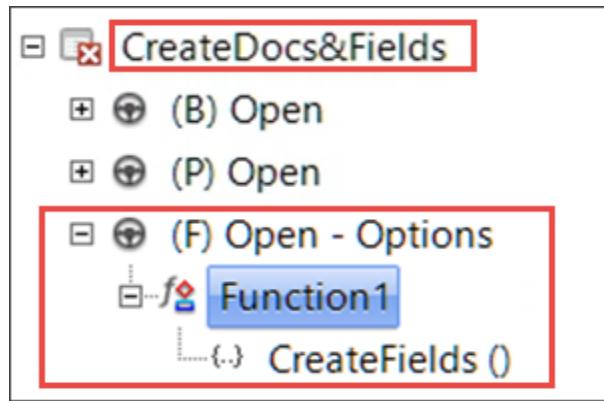
Figure 8-7. Define document hierarchy for an OMR field

The screen capture shows the Document hierarchy pane in Datacap Studio Rulemanager tab. On this pane, you can configure the Datacap objects (Document, Page, and Field) objects based on your business needs.

The following subfields are defined for an OMR field:

- Navigation
- ChildSeat
- FuelService

Configure the CreateDocs&Fields ruleset



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-8. Configure the CreateDocs&Fields ruleset

The screen capture shows the CreateDocs&Fields ruleset in Datacap Studio Rulemanager tab. The field level rule ((F) Open - Options) is defined for the OMR field to create fields.



Configure the Recognize ruleset and zones for the field

The left screenshot shows the Datacap Studio Rulemanager tab. A tree view displays a 'Recognize' ruleset with three items: '(B) Open', '(P) Open', and '(F) Open - Options'. The '(F) Open - Options' item is selected and expanded, showing a 'Configure OMR' function. This function contains two actions: 'rrSet ("0", "@F.OMRType")' and 'RecognizeFieldOCR_A ()'. The right screenshot shows a 'Rental Agreement' interface with an 'Options' section. It includes three checkboxes: 'GPS Navigation' (checked), 'Child Seat' (unchecked), and 'Fuel Service' (checked).

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-9. Configure the Recognize ruleset and zones for the field

The upper screen capture shows the Recognize ruleset in Datacap Studio Rulemanager tab. The field level rule (F) Open - Options is defined for the OMR field. The rule has a function Configure OMR and it contains the following actions:

- rrSet ("0", "@F.OMRType") from the RuleRunnerLogic > Datacap.Libraries.RuleRunnerLogic.Actions action library.

The setting represents the Square background to read non-dropout checkbox.

- RecognizeFieldOCR_A () from the RecognitionOCRA > RecognitionOCRA.OCRAActions action library.

This field-level action recognizes the zoned field and retrieves the fields value.

The lower screen capture shows the zones that are drawn for an OMR field with three checkboxes.

8.2. Identify the batch results output

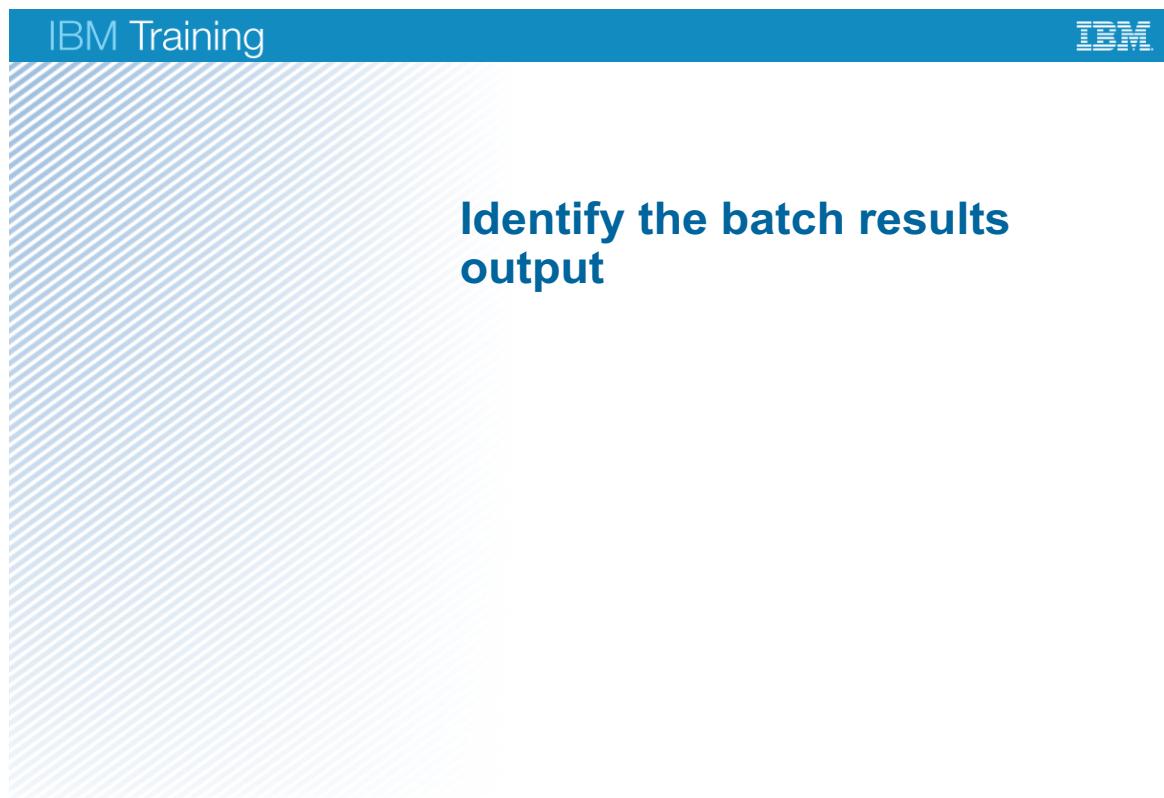
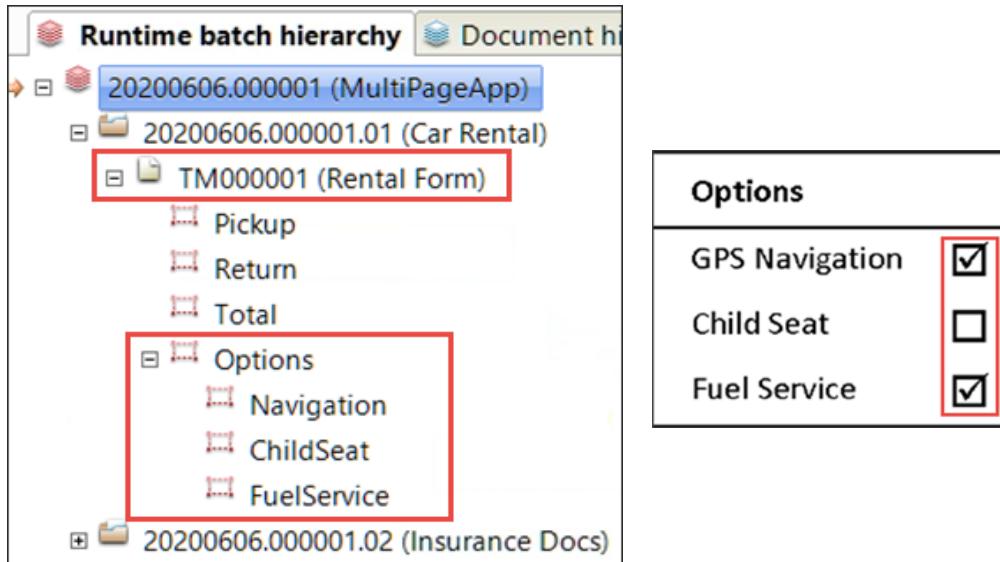


Figure 8-10. Identify the batch results output

Testing the application - Page ID task



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-11. Testing the application - Page ID task

The screen captures show the batch results in the Runtime batch hierarchy pane of the Datacap Studio Test view.

- The left screen capture shows the results for the PageID task. The Page (Rental Form) and Document (Car Rental) types are identified. The OMR fields are created (Navigation, ChildSeat, and FuelService for the Options field)
- The left screen capture shows the checkboxes on the actual form or image that is used for scanning for comparison.

Testing the application - Profiler task

The screenshot displays two windows side-by-side. On the left is the 'Runtime batch hierarchy' pane, which lists a tree structure of batches and their details. A red box highlights the 'Options' node under the '20200418.000004.01 (Car Rental)' batch, which contains three sub-options: Navigation, ChildSeat, and FuelService. To the right is a 'Options' configuration window showing three checkboxes: GPS Navigation (checked), Child Seat (unchecked), and Fuel Service (checked). The 'GPS Navigation' checkbox is also highlighted with a red box.

Options	
GPS Navigation	<input checked="" type="checkbox"/>
Child Seat	<input type="checkbox"/>
Fuel Service	<input checked="" type="checkbox"/>

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-12. Testing the application - Profiler task

The screen captures show the batch results in the Runtime batch hierarchy pane of the Datacap Studio Test view.

- The left screen capture shows the results for the Profiler task. The extracted values are shown for the OMR field. The values for the three check boxes are: 101 where the value 1 indicates that the checkbox is selected. Value 0 indicates that the checkbox is not selected.
- The right screen capture shows the check boxes on the actual form-image that is used for scanning. Compare the results of 101 with the selection of check boxes on the image.

Testing the application - Export task without a dictionary

```
C:\Datacap\MultiPageApp\export\20200418.000004.txt - Notepad++ [Administrator]
File Edit Search View Encoding Language Settings Tools Macro Run Plugins
20200418.000004.txt
1 ,Mon, Oct 4, 2010,Fri, Oct 8, 2010,$582.77,101,,
2 ,Matt Smith,12345786,CL-1328956
```

Figure 8-13. Testing the application - Export task without a dictionary

The screen capture shows the export text file in Notepad++. It contains two rows of comma-separated field values - one row for each page type.

After you complete the Export task for the batch, the file is created.

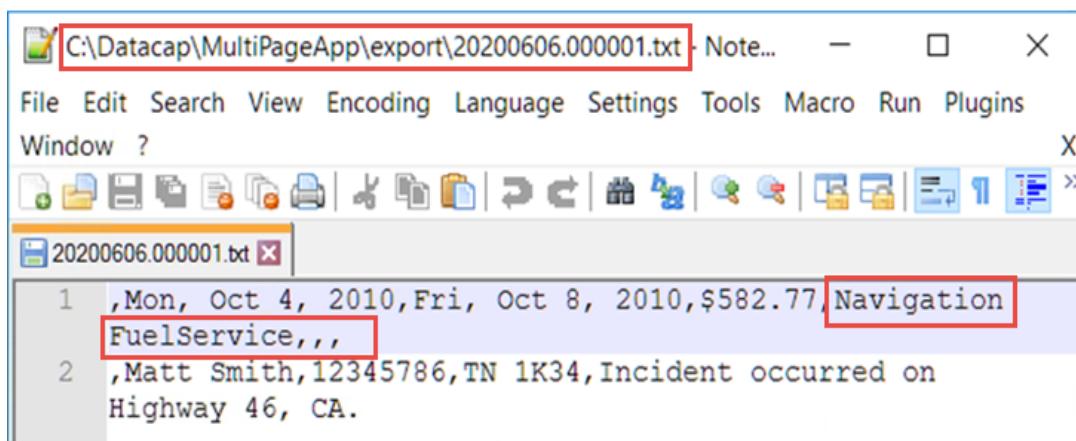
Recall the previous slide where the OMR results were shown on the Datacap Test tab and compare the values that are exported to the export file on this slide.

For the example on the slide, the Datacap application is not configured with a dictionary and the results are in ones and zeros.

Next slide shows the results for a Datacap application that is configured with a dictionary.

IBM Training 

Testing the application - Export task with a dictionary



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-14. Testing the application - Export task with a dictionary

The screen capture shows the export text file in Notepad++. It contains two rows of comma-separated field values - one row for each page type.

After you complete the Export task for the batch, the file is created.

Recall the previous slide where the OMR results were shown for a Datacap application that is not configured with a dictionary. The values are in ones and zeros.

Compare the previous values with the values on this slide.

For the example Datacap application on the slide, a dictionary is configured. The values are shown as text that you defined in the dictionary.

Unit summary

- Learn how to configure the application for recognition with OMR
- Identify the batch results output

Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-15. Unit summary

Review questions

1. You configure a field in Document hierarchy for OMR recognition and define a variable with the name MultiPunch and value 1. What does the variable indicate?

2. For a field in Document hierarchy, what does the variable RecogType=4 indicate?
 - A. OCR
 - B. Fingerprint
 - C. OMR
 - D. Barcode



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-16. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. You configure a field in Document hierarchy for OMR recognition and define a variable with the name MultiPunch and value 1. What does the variable indicate?

The answer is: [MultiPunch: 1 indicates that multiple selections are allowed for the field.](#)

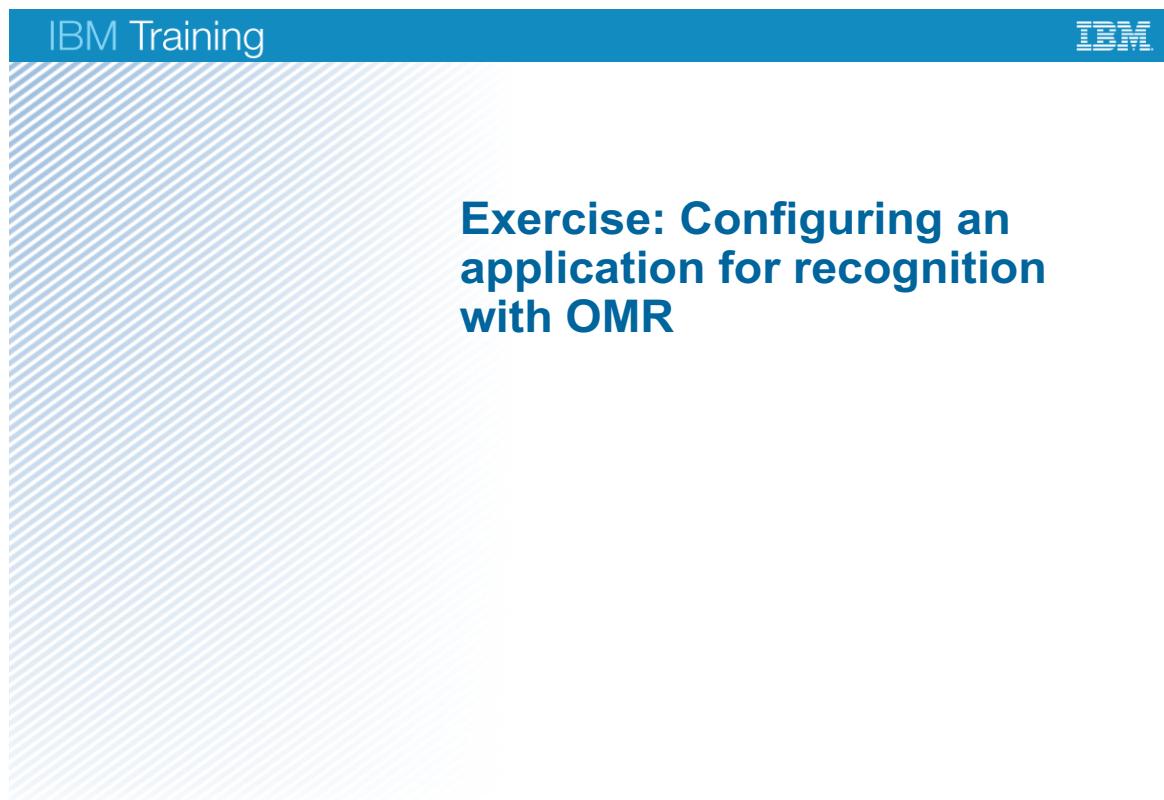


2. For a field in Document hierarchy, what does the variable RecogType=4 indicate?
 - A. OCR
 - B. Fingerprint
 - C. [OMR](#)
 - D. Barcode

The answer is [C.](#)

Figure 8-17. Review answers

- RecogType: 4 This value specifies the code for the recognition engine to use to read data from the field. The OMR (checkbox) fields require RecogType=4.



Configuring an application for recognition with OMR

© Copyright IBM Corporation 2020

Figure 8-18. Exercise: Configuring an application for recognition with OMR

Exercise objectives

- Configure OMR for a Datacap application
- Test the application



Figure 8-19. Exercise objectives

Unit 9. Exporting data to a FileNet Content Manager repository

Estimated time

00:30

Overview

This unit teaches how to configure the Datacap application to export scanned documents and data to a FileNet Content Manager repository.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Learn how to export data and documents to a FileNet Content Manager repository
- Describe how to verify the exported documents

Figure 9-1. Unit objectives

Topics

- Export data to a FileNet Content Manager repository
- Verify the exported documents

Figure 9-2. Topics

9.1. Export data to a FileNet Content Manager repository

Export data to a FileNet Content Manager repository

Exporting data to a FileNet Content Manager repository

© Copyright IBM Corporation 2020

Figure 9-3. Export data to a FileNet Content Manager repository

Export of scanned documents and data

- What does Export mean?
 - After capture processing is complete, send documents and index parameters to a data repository or the next business process
- You can send the data to several systems
 - Content Management Systems
 - Business-specific export options (LOB)
 - Case and Business Process Management Workflows
 - Databases and Electronic Data Interchange services

Figure 9-4. Export of scanned documents and data

The Export step is the last step of the capture process. When Capture processing is complete, you send data to a file or send documents and index parameters to a data repository or the next business process.

The following systems are some of the examples where the captured data is exported.

IBM Content Management systems

- IBM FileNet Content Manager

Use the actions in the FileNetP8 library to export data to a FileNet Content Manager repository

- IBM Content Manager

Use the actions in the IBMCMEExtended library to upload documents into an IBM Content Manager repository

Third-party Content Manager systems

- Microsoft SharePoint

Use the actions in the SPExport library to upload documents to a Microsoft SharePoint library.

- Any CMIS repository

Content Management Interoperability Services (CMIS) is an open standard

Use the actions in the CMISClient library to communicate with compliant content management systems

Business-specific export options

- Line of Business (LOB) applications
- Case-based Applications
- Business Process Management solutions
- Electronic Data Interchange (EDI)

Other export options

- Relational databases

Use the actions in the ExportToDatabase library to set up and write information to an export database

- XML export

Datacap can export data to an XML file by using the actions in the ExportToXML action library

- Local or network text file



Preparation for export to a FileNet repository: Document

The screenshot displays the IBM Administrative Console for Content Platform Engine. On the left, the Object Store: Sales tree view shows various categories like Administrative, Browse, Data Design, and Classes. The Data Design and Document nodes under Classes are highlighted with red boxes. On the right, a detailed view of the Car_Rental class definition is shown. The 'Property Definitions' tab is selected, and three properties are listed in a table:

	Property	Data Type
<input type="checkbox"/>	Pickup Date	Date Time
<input type="checkbox"/>	Return Date	Date Time
<input type="checkbox"/>	Total Cost	Float

Exporting data to a FileNet Content Manager repository

© Copyright IBM Corporation 2020

Figure 9-5. Preparation for export to a FileNet repository: Document

Before you export data to a FileNet Content Manager repository, you need to complete the configuration on the repository.

- Define a document class for each Datacap document type
- Define properties for the document class for each Datacap field that is extracted from the captured document
- Create a folder structure for exported documents

The screen capture shows the IBM Administrative Console for Content Platform Engine (ACCE), an administration client for the FileNet Content Manager.

In this example:

- Sales repository is open
- Car_Rental is the document class that is used for receiving the Datacap documents (Car Rental document type and Rental Form page type)
- Three properties (Pickup Date, Return Date, and Total Cost) are defined. The three properties are mapped to the three Datacap fields (Pickup, Return, and Total). The extracted values are stored in the Document properties on the repository.

The screenshot shows the IBM Datacap Navigator interface. On the left, there's a navigation pane titled "Datacap Navigator" with a "Browse" button highlighted by a red box. Below it is a tree view of a "Sales" folder containing various sub-folders like CBR_Docs, DCExport (which is also highlighted by a red box), Datacap Export, EDU Docs, Edit Service, Entry Templates, Items_For_Syncing, Orders, Pictures, Products, Role-based Redactions, Sample Docs, and Saved Searches. In the center, the path "Sales > DCExport" is shown above a list of documents. One document, "20200421.000001.01", is selected and highlighted by a red box. At the top of this list are columns for "Name", "Size", and "Modified By". To the right of the list is a "Properties" panel with sections for "Class", "Document Title", "Pickup Date", "Return Date", and "Total Cost". The "Class" field is set to "Car_Rental". The "Document Title" field contains the value "20200421.000001.01". The "Pickup Date" is listed as "10/3/2010, 8:00 PM". The "Return Date" is listed as "10/7/2010, 8:00 PM". The "Total Cost" is listed as "582.77". At the bottom of the properties panel, it says "Added By: p8admin". The top right corner of the interface has an "admin" icon and a help (?) icon.

Figure 9-6. Preparation for export to a FileNet repository: Folder

The previous slide showed the configuration that is required for the document class. This screen capture shows the IBM Content Navigator Browse feature to view the documents and folders from the FileNet Content Manager repository.

- On the left pane, a folder (DCExport) is set up for the export documents
- In the middle pane, a document is listed which is exported from Datacap.
- On the right pane, properties for the document are shown. The properties contain the data values that are extracted from the Datacap fields.



Export to FileNet P8 ruleset: Set parameters

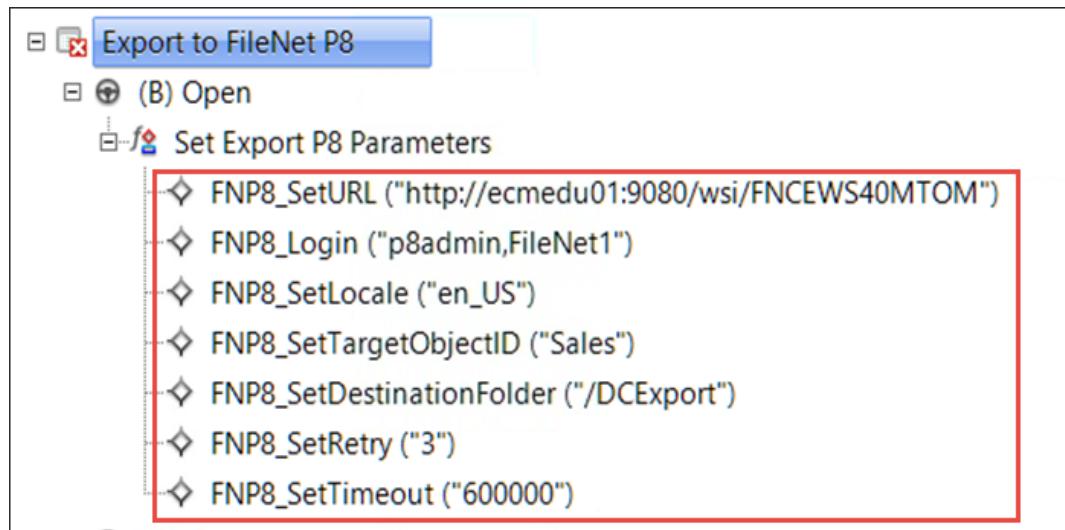


Figure 9-7. Export to FileNet P8 ruleset: Set parameters

The screen capture shows the ruleset for exporting the documents and data to a FileNet (P8) Content Manager repository.

The following list provides the actions and their description:

- FNP8_SetURL: Sets the URL for the FileNet P8 Web Service
- FNP8_Login: Sets the user ID and password to log in to the FileNet (P8) Content Manager system
- FNP8_SetLocale: Sets the locale, for example: en_US
- FNP8_SetTargetObjectID: Sets the name of the FileNet P8 repository (known as Object store) for the document upload
- FNP8_SetDestinationFolder: Sets the destination folder for the documents that are uploaded
- FNP8_SetRetry: The number of attempts it makes to upload
- FNP8_SetTimeout: Sets the timeout for the FileNet P8 Web Service

The image contains two side-by-side screenshots of the IBM Training interface, both titled "Export: Set document parameters and map property values".

Upper Screenshot: Shows the rule (D) Open selected. It includes three actions under "Set Doc Parameters": FNP8_SetDocClassId ("Car_Rental"), FNP8_SetDocTitle ("@ID"), and FNP8_SetFileType ("pdf"). These actions are highlighted with a red box.

```

graph TD
    A[Export to FileNet P8] --> B((B) Open)
    A --> C((D) Open)
    C --> D[Set Doc Parameters]
    D --> E[FNP8_SetDocClassId ("Car_Rental")]
    D --> F[FNP8_SetDocTitle ("@ID")]
    D --> G[FNP8_SetFileType ("pdf")]
  
```

Lower Screenshot: Shows the rule (P) Open selected. It includes three actions under "Map Field Values": FNP8_SetProperty ("PickupDate,@P\Pickup,DateTime"), FNP8_SetProperty ("ReturnDate,@P\Return,DateTime"), and FNP8_SetProperty ("TotalCost,@P\Total,Float"). These actions are highlighted with a red box.

```

graph TD
    A[Export to FileNet P8] --> B((B) Open)
    A --> C((D) Open)
    A --> D((P) Open)
    D --> E[Map Field Values]
    E --> F[FNP8_SetProperty ("PickupDate,@P\Pickup,DateTime")]
    E --> G[FNP8_SetProperty ("ReturnDate,@P\Return,DateTime")]
    E --> H[FNP8_SetProperty ("TotalCost,@P\Total,Float")]
  
```

Exporting data to a FileNet Content Manager repository © Copyright IBM Corporation 2020

Figure 9-8. Export: Set document parameters and map property values

The screen captures show two rules of the Export to FileNet P8 ruleset.

The upper screen capture shows the (D) Open rule and it is run at the document level. The actions define the following settings:

- FNP8_SetDocClassId: Identifies the Document class in the FileNet (P8) Content Manager repository to set for the document that is exported (Example: Car_Rental)
- FNP8_SetDocTitle: Sets the title for the document that is uploaded with the value specified (Example: Batch ID value)
- FNP8_SetFileType: Sets the file type for the document that is uploaded (Example: pdf)

The lower screen capture shows the (P) Open rule and it is run at the page level. The actions map the Datacap fields with the document properties in the repository. It also specifies what is the data type for the incoming value.

Export to FileNet P8 ruleset: Document upload

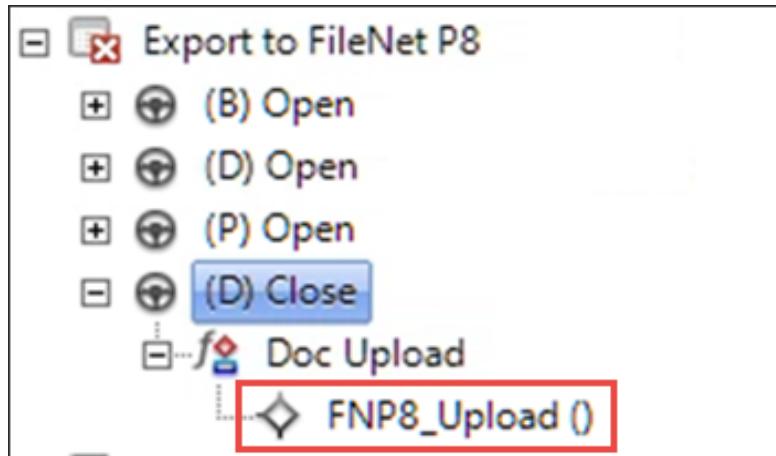
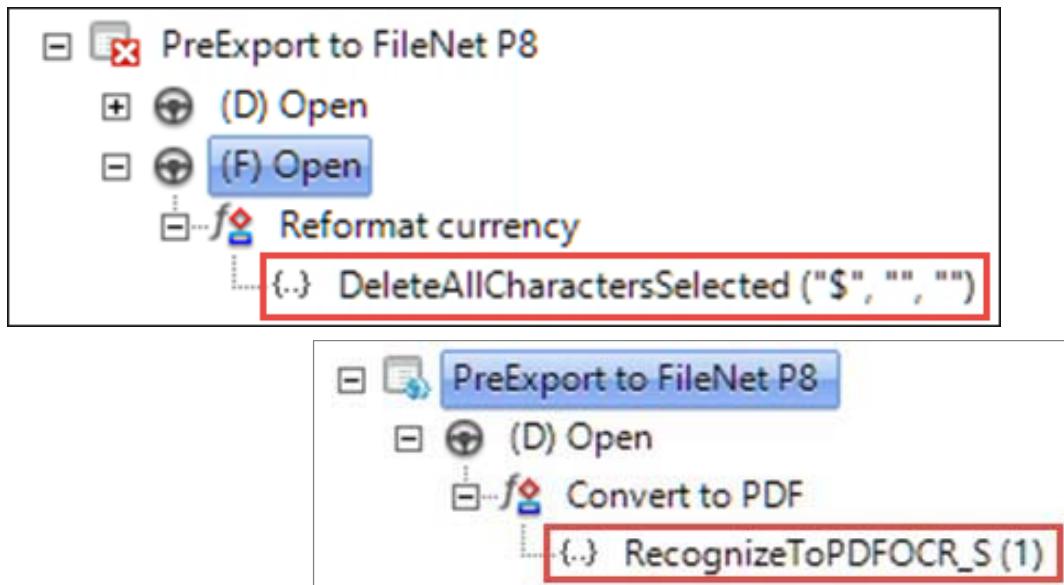


Figure 9-9. Export to FileNet P8 ruleset: Document upload

After you set all the required parameters, you call the action FNP8_Upload() to upload the document.

Preparing the data for the repository upload



Exporting data to a FileNet Content Manager repository

© Copyright IBM Corporation 2020

Figure 9-10. Preparing the data for the repository upload

The screen captures show the ruleset that prepares the data to export to the FileNet (P8) Content Manager repository.

- The upper screen capture shows the (F) Open rule and it is run at the field level.

The extracted data in Datacap might not be compatible to export to the repository and you need to clean the data. In this example, the extracted total cost contains a \$ (dollar) symbol along with a currency value. You need to remove the \$ symbol to export the value as float data type.

- The lower screen capture shows the (D) Open rule and it is run at the document level.

In the example, the original image files that are used for scanning are of TIFF file type. You want to upload these documents as PDF file type. You need to add an action (RecognizeToPDFOCR_S) that converts the TIFF file to PDF.

9.2. Verify the exported documents

Verify the exported documents

Exporting data to a FileNet Content Manager repository

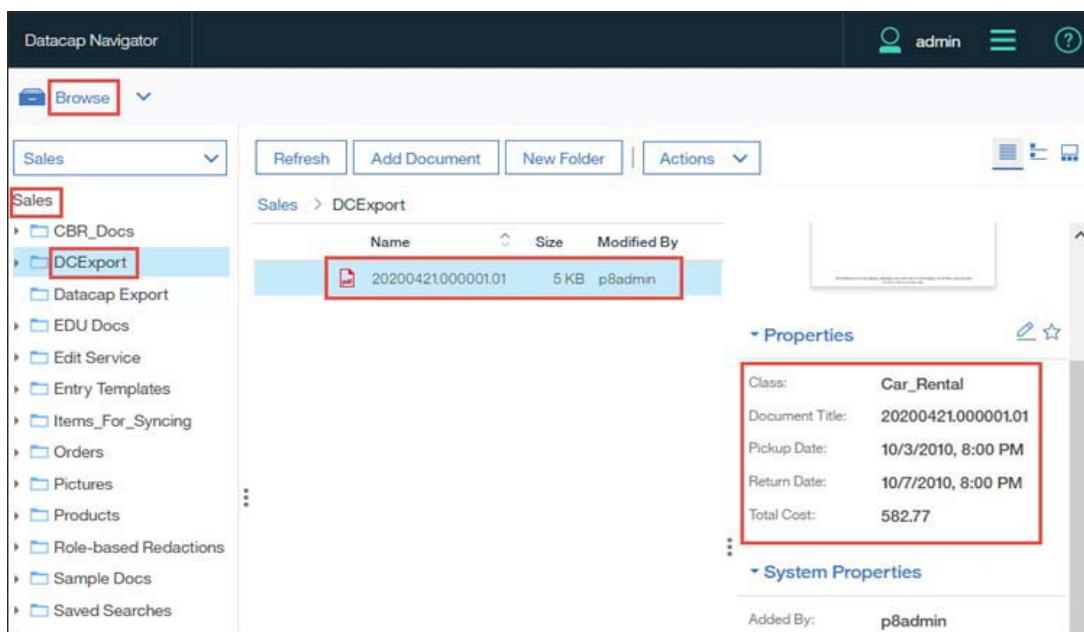
© Copyright IBM Corporation 2020

Figure 9-11. Verify the exported documents

IBM Training



Verifying the export on the repository



The screenshot shows the IBM Datacap Navigator interface. On the left, the 'Sales' folder is expanded, and the 'DCExport' folder is selected. In the middle pane, a document named '20200421.000001.01' is listed with a size of 5 KB and modified by 'p8admin'. On the right, the 'Properties' panel displays the following details:

Class:	Car_Rental
Document Title:	20200421.000001.01
Pickup Date:	10/3/2010, 8:00 PM
Return Date:	10/7/2010, 8:00 PM
Total Cost:	582.77

Below the properties, the 'System Properties' section shows 'Added By: p8admin'.

Exporting data to a FileNet Content Manager repository

© Copyright IBM Corporation 2020

Figure 9-12. Verifying the export on the repository

This screen capture shows the IBM Content Navigator Browse page to view the documents and folders from the FileNet Content Manager repository.

- On the left pane, a folder (DCExport) is set up for the export documents
- In the middle pane, a document is listed which is exported from Datacap.
- On the right pane, properties for the document are shown and it has values that are extracted from the Datacap fields.

Unit summary

- Learn how to export data and documents to a FileNet Content Manager repository
- Describe how to verify the exported documents

Figure 9-13. Unit summary

Review questions

1. True or False: After capture processing is complete, you can send data to a text file, a repository, or the next business process.
2. To configure the export to a FileNet P8 repository, you add the FNP8_SetTargetObjectID action. Which object name must you set as the parameter value for this action?
 - A. Document
 - B. Repository
 - C. Folder
 - D. Property



Figure 9-14. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. **True** or False: After capture processing is complete, you can send data to a text file, a repository, or the next business process.
The answer is True.

2. To configure the export to a FileNet P8 repository, you add the FNP8_SetTargetObjectID action. Which object name must you set as the parameter value for this action?
 - A. Document
 - B. **Repository**
 - C. Folder
 - D. Property**The answer is B.**



Figure 9-15. Review answers

Exercise: Exporting data to a FileNet Content Manager repository

Exporting data to a FileNet Content Manager repository

© Copyright IBM Corporation 2020

Figure 9-16. Exercise: Exporting data to a FileNet Content Manager repository

Exercise objectives

- Export data and documents to a FileNet Content Manager repository
- Verify the exported documents



Figure 9-17. Exercise objectives

Unit 10. Creating page layouts

Estimated time

00:20

Overview

In this unit, you learn how to create page layouts. The layout file groups text into blocks as a person would look at the document. Each block might have the default type of block or a specific type such as title or table.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacaptoc.doc/datacap_9.1.7.htm

Unit objectives

- Describe Datacap Insight Edition concepts
- Learn how to create page layouts

[Creating page layouts](#)

© Copyright IBM Corporation 2020

Figure 10-1. Unit objectives

Topics

- Introduction to Datacap Insight Edition
- Creating page layouts

[Creating page layouts](#)

© Copyright IBM Corporation 2020

Figure 10-2. Topics

10.1. Introduction to Datacap Insight Edition

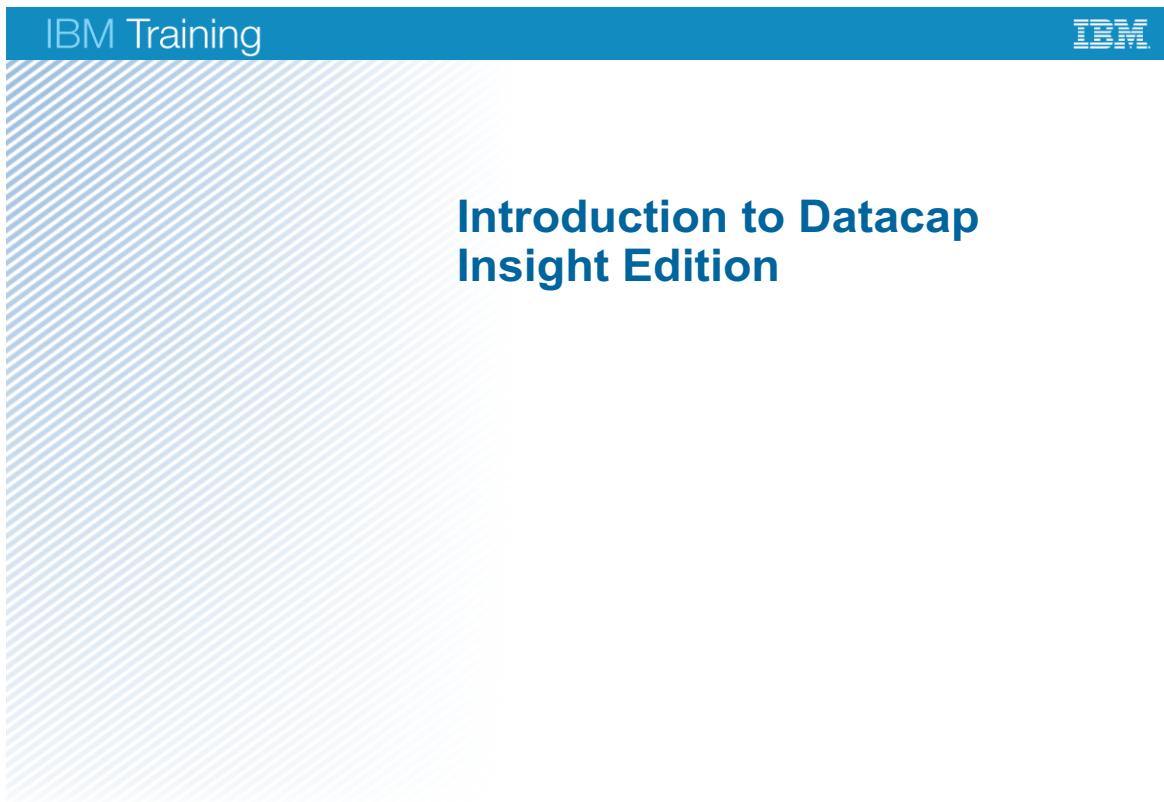


Figure 10-3. Introduction to Datacap Insight Edition



Unstructured documents

- All documents do not have a predictable format, for example:
 - Letters, Contracts
- Datacap solution needs to be customized
- Datacap offers Insight edition tools to handle unstructured documents

The image shows a collage of several unstructured document examples:

- BILLS OF LADING:** A form with fields for Consignee, Street, Description, City/State/Zip, Name, Company, Email, and Phone.
- DRAFTS:** A form titled "DRAFTS" with sections for "BILL OF EXCHANGE" and "CERTIFICATES".
- CERTIFICATES:** A certificate for "CERTIFICATE OF QUALITY ANALYSIS" dated 31-03-2015, issued by "SAPLPLR-11403040-H15".
- BILLS OF EXCHANGE:** A bill of exchange with fields for Date, AMT TO, ILL OR, and AN BLA.
- LEASES:** A lease agreement document.
- CONTRACTS:** A contract for photography services.
- PACKING LISTS:** A packing list from "Globe Vocational Enterprises Limited (GVE) Ltd.".

Creating page layouts

© Copyright IBM Corporation 2020

Figure 10-4. Unstructured documents

The screen captures show examples of unstructured forms.

Most of the important information that an organization receives is not on a form but is in a letter, in an email message, or in a PDF file. And traditional document capture systems just can't handle that sort of unstructured data.

Cognitive data extraction that uses document structure

- Understands the document structure in addition to the text content: Blocks, sections, tables, paragraph, labels.
- Extracted rich text data

IBM Datacap uses cognitive technologies such as Natural Language Processing to help change how one approaches and understands the unstructured documents. It can ingest this unstructured data, understand it, reason about it, and learn from it. It can look at the structure and content to determine a document type. It can use text analytics to identify entities, patterns, and keywords to extract important information.

Cognitive capture automates the process of document transactions by using a combination of advanced document imaging, Content Analytics, and machine learning techniques to process all document types.

In this course, solutions for structured and unstructured documents are discussed.

IBM Datacap Insight Edition

- A cognitive capture solution that combines advanced imaging, natural language processing, and machine learning technologies
- Automates the processing of unstructured documents through multi-level analysis
- Distinguishes between different types of documents based on their structure.
- Applies reasoning, logic, and context-sensitive analysis to identify and classify information.
- Uses a new structural layout

Figure 10-5. IBM Datacap Insight Edition

In the past, getting data from documents involved pre-defining templates or writing complex regular expression or search rules. But with Datacap Insights Edition, you can extract data more easily.

Datacap Insight Edition is an add-on to Datacap that includes new Datacap actions for cognitive capabilities and data analytics.

Cognitive capture is ideal for a document when you don't know the format of the document or the location of fields. Examples include contracts, correspondent documents such as legal documents, claims, and letters.

A key feature in the Datacap Insight application is that data can be identified and extracted before the pages are identified or classified. This feature allows the information in the document to be used for identification and classification.

10.2. Creating page layouts

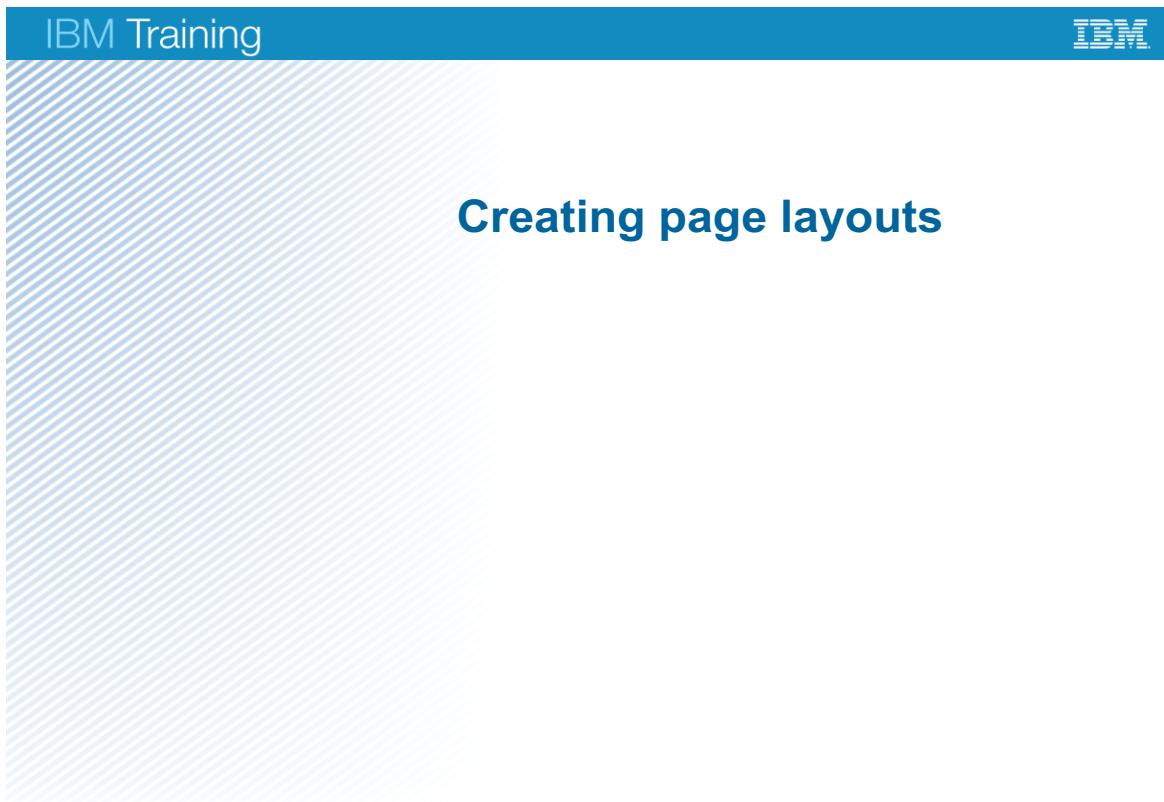
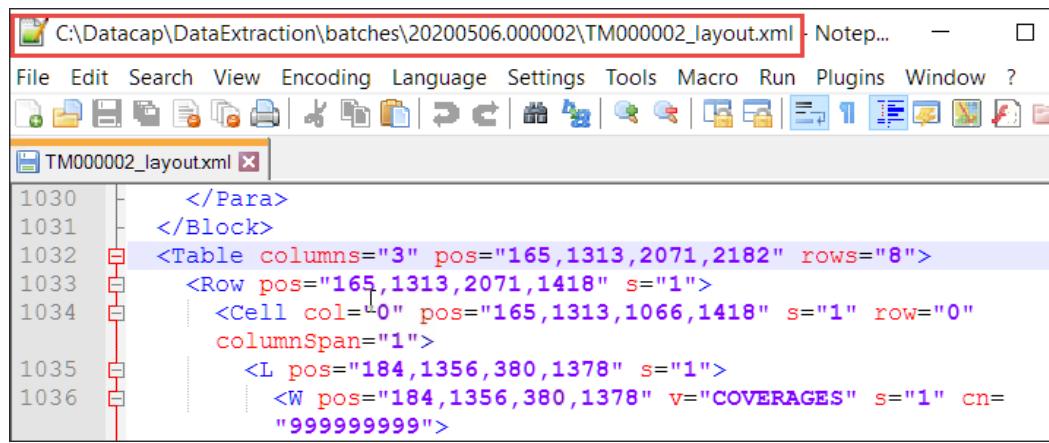


Figure 10-6. Creating page layouts



```

C:\Datacap\DataExtraction\ batches\20200506.000002\TM000002_layout.xml - Notep... — □
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
TM000002_layout.xml x
1030      </Para>
1031      </Block>
1032      <Table columns="3" pos="165,1313,2071,2182" rows="8">
1033        <Row pos="165,1313,2071,1418" s="1">
1034          <Cell col="0" pos="165,1313,1066,1418" s="1" row="0"
               columnSpan="1">
1035            <L pos="184,1356,380,1378" s="1">
1036              <W pos="184,1356,380,1378" v="COVERAGES" s="1" cn=
               "999999999">

```

Figure 10-7. Page layout XML file

The screen capture shows a sample page layout XML file. Data can be extracted from the layout file by using Datacap actions.

The PDF and image files are processed to produce a layout XML structure for each page. Full page OCR recognition or PDF conversion actions create the layout XML structure. More actions can be used to enhance the internal format of the layout and supplement it with data from text analytics.

The layout file contains the results of recognition. The layout file groups text into blocks as a person would look at the document. Each block might have the default type of block or a specific type such as title or table.

Following are the examples of types of elements that might be present in the layout XML file:

- Block
- Header and Footer
- Title
- Heading
- Barcode
- Table, Row, and Cell
- Para

The PDFFREDocumentToImage action

- bool PDFFREDocumentToImage (string resolution, string compressionBW, string compressionColor, string compressionGray, string extensionBW, string extensionColor, string extensionGray, string convertMode, string useFastBinarization, string jpegQuality)

The screenshot shows the IBM Action library interface. On the left, there is a tree view under the 'Convert' category. One item, 'PDFFREDocumentToImage', is highlighted with a red box. On the right, a table titled 'Parameters' lists the parameters for this action with their sample values.

Parameters	
string resolution	300
string compressionBW	18
string compressionColor	33
string compressionGray	32
string extensionBW	.tif
string extensionColor	.tif
string extensionGray	.tif
string convertMode	0
string useFastBinarization	false
string ipeaQuality	100

Figure 10-8. The PDFFREDocumentToImage action

The left screen capture shows the PDFFREDocumentToImage action in the Action library list. The right screen capture shows the parameters with sample values for the action.

This action runs recognition in addition to text extraction. Before you call the action, enable this capability by setting the DCO variable *y_createLayout* to "1". By default, this feature is turned off.

When this option is turned on, a layout xml file (for example tm000001_layout.xml) is created per image that is extracted.

The PDFFREDocumentToImage action converts a PDF file to TIFF format so that you can use standard Datacap methods to further process and verify documents and data.

- The convertMode parameter value 0 (that is used for this activity) preserves the color. The extension that is used is TIF.

The Recognize action

- Performs document layout analysis and OCR and also generates a layout XML file
 - Example: TM000001_layout.xml
- The layout XML file:
 - Groups text into blocks
 - Retains font and color attributes
- This action returns False if the ruleset with this action is not bound to a page object of the document hierarchy

Figure 10-9. The Recognize action

The layout file groups text into blocks similar to how a person would see and identify the structure in the document. For example, a page can have items such as tables, paragraphs, and lines, which are all a type of a block.

The layout XML file retains font and color attributes (saved in CSS format) for the text that is used for extracting data and reconstructing the document in a new format.

Recognizing the PDF and the Image files

- PDFs that contain machine printable text
 - Use actions that recognize the PDF simultaneously as it creates a TIF image
- PDFs that contain scans of forms
 - Convert the PDF to images and perform recognition directly on the newly generated page images

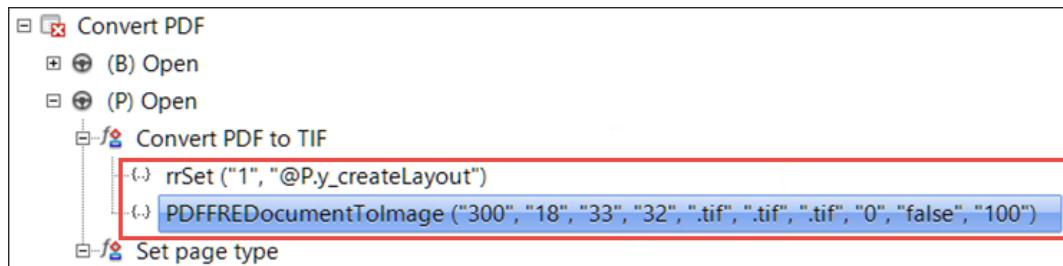


Figure 10-10. Recognizing the PDF and the Image files

The screen capture shows a sample ruleset that contains the PDFFREDocumentToImage action and the rrSet action for setting the DCO variable *y_createLayout* to "1". These two actions together recognize the PDF and it creates a TIF image.

By enabling the creation of a layout file, the engine recognizes the PDF and creates a TIF image for each page in the PDF.

You can use this method for PDFs that contain printable text that is generated directly from office type applications such as a word processor.

For other PDF documents, you can convert the PDF to images first and then run the recognition on the newly generated page images.

- With this approach, you can manipulate the documents before recognition with image enhancements.
- After the image quality is improved, you can do the recognition to provide better recognition results compared to directly recognizing the PDF.

Analyze layout

- AnalyzeLayout enhances and refines the layout XML file
- CreateCcoFromLayout creates CCO file from the layout XML
- Creates a text file for the page with the recognized values
- Creates an HTML document based on the text and the layout XML file

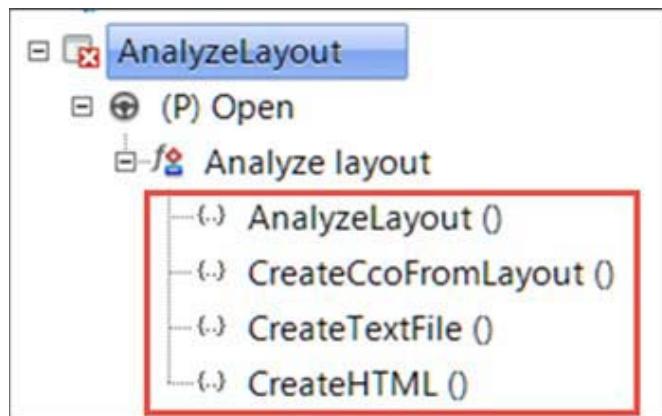


Figure 10-11. Analyze layout

The screen capture shows a sample ruleset that contains the following actions:

- AnalyzeLayout
 - It is from the Document Analytics > Document Analytics.Actions action library
 - It enhances and refines the layout XML and improves the labeling and grouping of text blocks that are identified
 - It requires a previously created layout file (for example: TM000001_layout.xml)
- CreateCcoFromLayout
 - It is from the SharedRecognitionTools action library
 - It transfers the characters, confidence, lines, and words from the layout XML to Datacap CCO format and creates a CCO file for each page in the batch.
 - The CCO file that the action creates is required to use the Locate actions and perform click and key during verification
- CreateTextFile()
 - It is from the SharedRecognitionTools action library
 - It creates a text file for the current page and adds the recognized values to page.

- It is helpful for debugging purposes and you can check what recognition values are placed in the CCO file.
- CreateHTML()
 - It is from the Document Analytics action library
 - It creates an HTML document based on the text and format that is captured in the layout XML file.
 - This action is helpful where an HTML document is needed for processing or viewing the document in another product after capture is complete.
 - It is an optional action.

Populate Layout field in DCO

- CopyAllBlocks copies the fields to the batch document hierarchy
- SetDCOStatus hides the layout fields in user interfaces

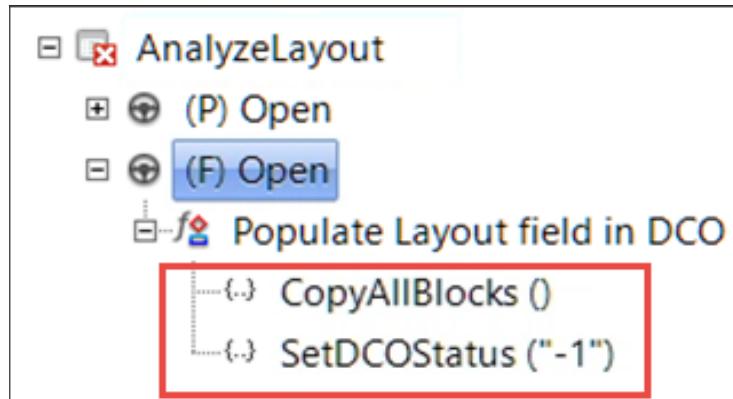


Figure 10-12. Populate Layout field in DCO

The screen capture shows a sample ruleset that contains the following actions:

- CopyAllBlocks
 - It copies the fields (as subfields to Layout) to the batch document hierarchy for each block of text.
 - It helps visualize the blocks within Datacap Studio.
- SetDCOStatus
 - It helps to hide the layout fields in user interfaces. The layout fields are not intended to be used from the standard user interfaces.
 - It is also a good practice to delete the Layout subfields by using the DeleteFields action when the development and testing are complete.



Creating page layouts

© Copyright IBM Corporation 2020

Figure 10-13. Test the application

The left screen capture shows the batch process results for the Page ID task in the Datacap Studio Test tab. The right screen capture shows the actual document that is used for the scanning.

For the application, when you run the Page ID task, the Layout XML file is created. The CopyAllBlocks action copies the fields as subfields to Layout in the batch document hierarchy for each block of text. The field values are shown as green text. Compare the Title block in the results with the title of the document.

Unit summary

- Describe Datacap Insight Edition concepts
- Learn how to create page layouts

[Creating page layouts](#)

© Copyright IBM Corporation 2020

Figure 10-14. Unit summary

Review questions

1. True or False: The Recognize action creates the layout XML file and groups the text from the scanned document into various types of blocks
2. What does the AnalyzeLayout action do?
 - A. It analyzes and removes the duplicate sentiments from the layout XML file
 - B. It analyzes and removes the duplicate keywords from the layout XML file
 - C. [CORRECT] It fine-tunes the layout XML file and improves the labeling and grouping of text blocks that are identified
 - D. It fine-tunes the layout XML file to verify the alphabetical order of the text



Figure 10-15. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. True or False: The Recognize action creates the layout XML file and groups the text from the scanned document into various types of blocks.

The answer is True.



2. What does the AnalyzeLayout action do?

- A. It analyzes and removes the duplicate sentiments from the layout XML file
- B. It analyzes and removes the duplicate keywords from the layout XML file
- C. It fine-tunes the layout XML file and improves the labeling and grouping of text blocks that are identified
- D. It fine-tunes the layout XML file to verify the alphabetical order of the text

The answer is C.

Figure 10-16. Review answers

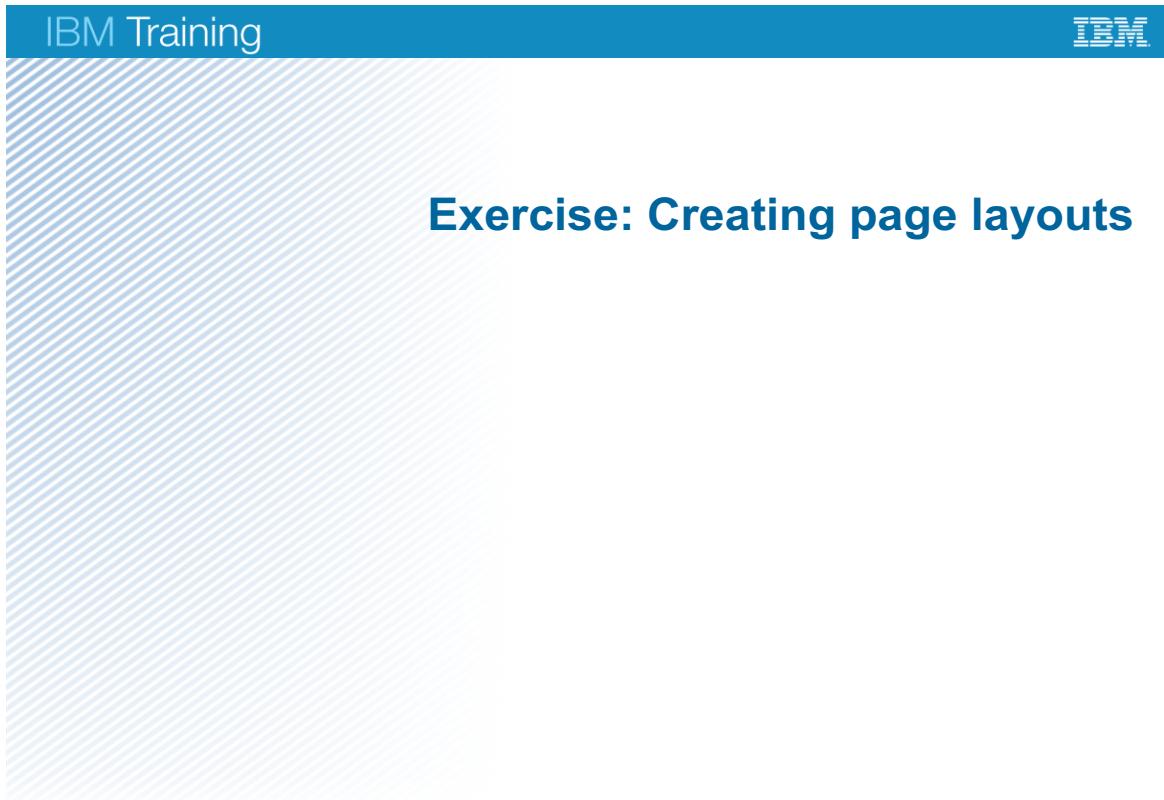


Figure 10-17. Exercise: Creating page layouts

Exercise objectives

- Configure a Datacap application to create page layouts
- Test the application and verify the layout XML file



[Creating page layouts](#)

© Copyright IBM Corporation 2020

Figure 10-18. Exercise objectives

Unit 11. Creating a virtual page block and extracting data from tables

Estimated time

00:30

Overview

In this unit, you learn how to create a virtual page block and extract data from tables.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacapoc.doc/datacap_9.1.7.htm

Unit objectives

- Describe how to create a virtual page block
- Learn how to configure data extraction from tables

Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-1. Unit objectives

Topics

- Create a virtual page block
- Configure data extraction from tables

Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-2. Topics

11.1. Create a virtual page block

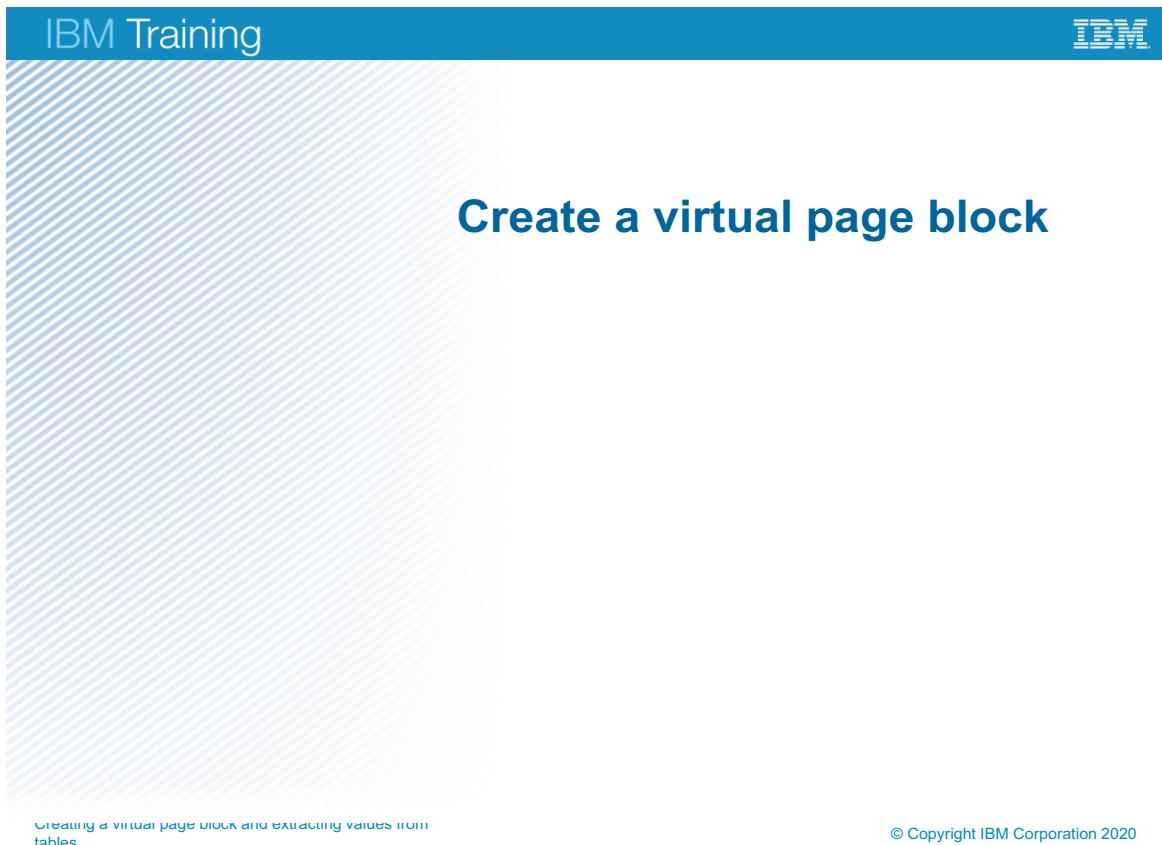


Figure 11-3. Create a virtual page block



What is a virtual page block (VPB)?

GENERAL INSURANCE COMPANY
VERIFICATION OF COVERAGE
(SEE BELOW UNDER CAUTIONARY NOTE)

INSURED
ROBERT HANSON
1234 Oak St.
McKinley, CA, 75070

To whom it may concern:
This letter is to verify that we have issued the policyholder coverage under the above policy number for the dates indicated in the effective and expiration date fields for the vehicle listed. This should serve as proof that the below mentioned vehicle meets or exceeds the financial responsibility requirement for your state.

This verification of coverage does not amend, extend or alter the coverage afforded by this policy.

Vehicle Year: 2015
Make: BMW
Model: 535i
VIN: WBSFH4H19F83241810

COVERAGES	LIMITS	DEDUCTIBLES
BODILY INJURY LIABILITY	\$100,000 / \$300,000	
PROPERTY DAMAGE LIABILITY	\$50,000	
UNINSURED & UNDERINSURED MOTORISTS	\$50,000 / 360,000	
COMPREHENSIVE		ACV LESS \$1,000
COLLISION		\$1000 DED/WAIVER
EMERGENCY ROAD SERVICE	FULL	NON-DED
RENTAL REIMBURSEMENT	\$35/DAY - \$1,050 MAX	

Lienholder Additional Insured Interested Party

BMW FINANCIAL SERVICES NA
P.O. BOX 3331
ATTENTION: INSURANCE
SAFETY TEAM, ALEX, 11347

Creating a virtual page block and extracting values from tables

© Copyright IBM Corporation 2020

Figure 11-4. What is a virtual page block (VPB)?

A virtual page block (VPB) is a new page that is a subset of an existing page. A VPB can be used to reduce the amount of data that needs to be processed as compared to its source file. You can run the Datacap actions on a VPB as you would any other page.

To create a VPB from a page, you need to identify keywords that indicate the start and end of a section of a page.

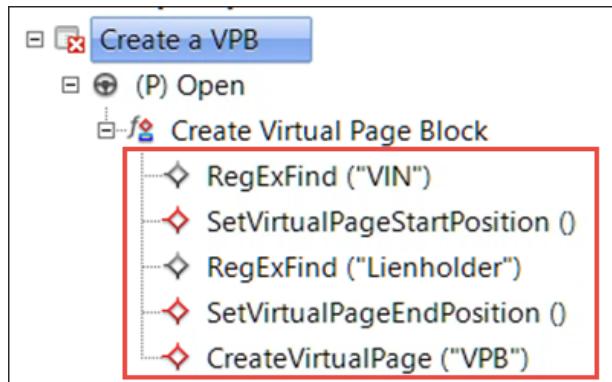
The screen capture shows a full page, and the proposed VPB is highlighted in red.

The VPB contains primarily the table in the page.

For the sample document that is used in this unit, the words *VIN* and *Lienholder* can be used as start and end. In the following sections, you use these words to identify the start and end of the block.

Sample ruleset to create a virtual page block (VPB)

- Use Locate rules to define the VPB to identify a word at the start and a word at the end of the block.



Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-5. Sample ruleset to create a virtual page block (VPB)

The screen capture shows a sample ruleset to create a virtual page block (VPB) that contains the following actions:

- The RegExFind action locates the first occurrence of a word or phrase (Example: VIN) that you passed as the parameter on the current page.
 - The search is case-sensitive.
 - The word or phrase is used to mark the start position
- After the action identifies the start position for the VPB, the SetVirtualPageStartPosition action sets it as the start position.
- You add the RegExFind action the second time to locate the first occurrence of a word or phrase (Lienholder) that you passed as the parameter on the current page.

The word or phrase is used to mark the end position.

- The SetVirtualPageEndPosition action establishes the end of the VPB.
- The CreateVirtualPage action creates the virtual page with the start and end that you defined.

The name, which you passed in the parameter for the virtual page (VPB), is already defined in the DCO.

Verifying a virtual page block (VPB)



1000		</I>
1001		<L pos="167,1228,626,1254" s="2">
1002		<W pos="167,1228,238,1254" v="VIN:" s="2" cn="9999">
1003		<C pos="167,1228,192,1254" v="V" s="2" cn="10" />
1004		<C pos="192,1228,204,1254" v="I" s="2" cn="10" />

Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-6. Verifying a virtual page block (VPB)

Test the rulesets for the VPB in Datacap Studio.

The upper screen capture shows the Runtime batch hierarchy pane in Datacap Studio Test tab. The VPB page that is created and its field Table are highlighted.

By using the page layout actions, you can create a layout XML for VPB. The lower screen capture shows the layout XML for the VPB.

In the example that is used for this unit, the scan page contains text and table. You create a VPB that contains only the table. The VPB reduces the amount of data that needs to be processed as compared to its source file. In the following slides, you learn how to extract the data from the table in the VPB.

11.2. Configure data extraction from tables

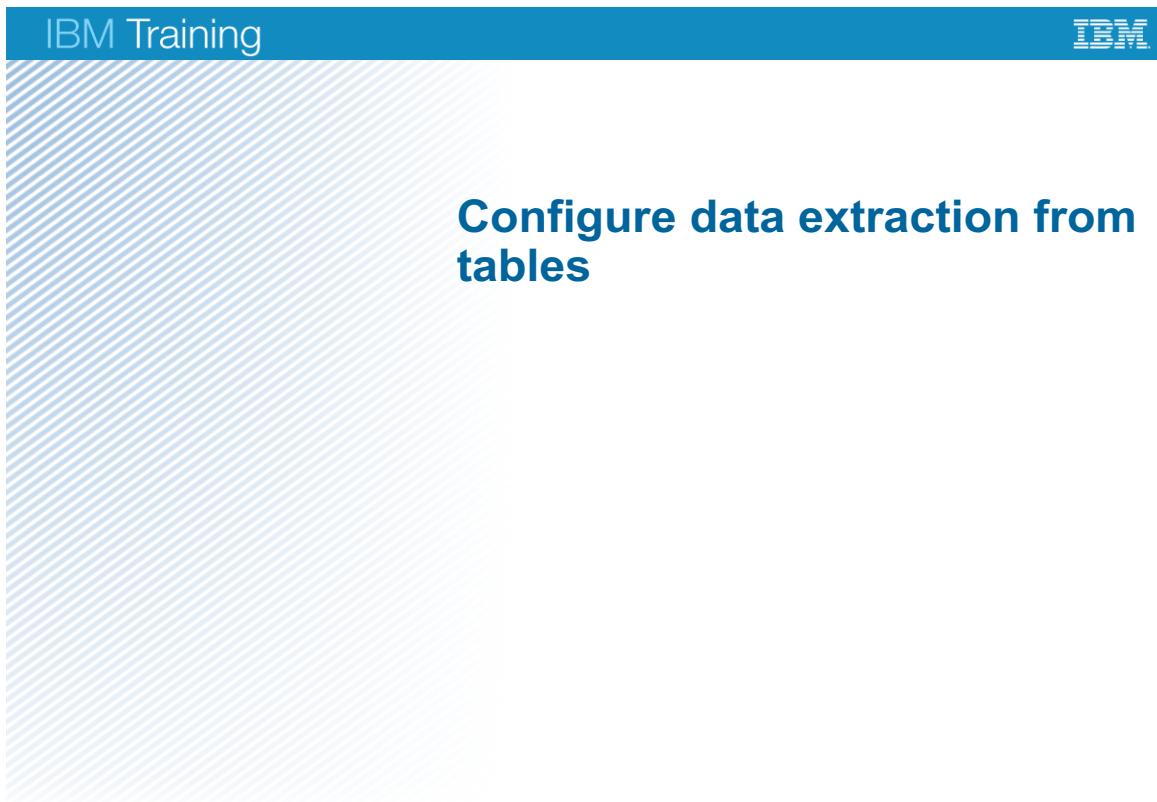
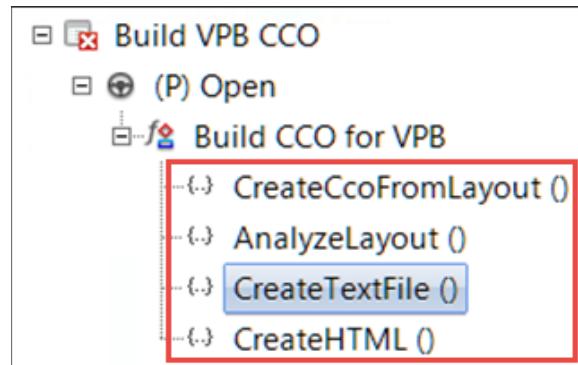


Figure 11-7. Configure data extraction from tables



Sample ruleset for Analyze layout for VPB

- The ruleset contains actions for refining the layout XML for VOPB.



Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-8. Sample ruleset for Analyze layout for VPB

In the previous unit, you learn about the actions on this slide.

The screen capture shows a sample ruleset that contains the following actions:

- AnalyzeLayout
 - It is from the Document Analytics > Document Analytics.Actions action library
 - It enhances and refines the layout XML and improves the labeling and grouping of text blocks that are identified
 - It requires a previously created layout file (for example: TM000001_layout.xml)
- CreateCcoFromLayout
 - It is from the SharedRecognitionTools action library
 - The action is required if you plan to use the Locate actions or click and key during verification
 - It transfers the characters, confidence, lines, and words from the layout XML to Datacap CCO format and creates a CCO file for each page in the batch.

- CreateTextFile()
 - It is from the SharedRecognitionTools action library
 - It creates a text file for the current page and adds the recognized values to page.
 - It is helpful for debugging purposes and you can check what recognition values are placed in the CCO file.
- CreateHTML()
 - It is from the Document Analytics action library
 - It creates an HTML document based on the text and format that is captured in the layout XML file.
 - This action is helpful where an HTML document is needed for processing or viewing the document in another product after capture is complete.
 - It is an optional action.

The FindTableValueRegEx action

The upper part of the screenshot shows a table from a scanned document. The table has two columns: 'COVERAGES' and 'LIMITS'. The rows contain the following data:

COVERAGES	LIMITS
BODILY INJURY LIABILITY	\$100,000 / \$300,000
PROPERTY DAMAGE LIABILITY	\$50,000
UNINSURED & UNDERINSURED MOTORISTS	\$30,000 / 360,000

The lower part of the screenshot shows the properties of the 'FindTableValueRegEx' action in a software interface. The 'Action' field is set to 'FindTableValueRegEx'. The 'Namespace' field is 'DocumentAnalytics.Actions'. The 'Object instance' field is empty. The 'Parameters' section shows the following values:

string FindHeadingValue	COVERAGES
string FindRowValue	BODILY INJURY LIABILITY
string QueryColumn	LIMITS
string RowName	Bodily Injury
string CellName	Limits
string ValueStoreLocation	Table
string FailMissingResults	False

Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-9. The *FindTableValueRegEx* action

The upper screen capture shows the table on the scanned document. The values that are used for the action parameters are highlighted.

The lower screen capture shows the parameter values for the *FindTableValueRegEx* action.

Observe the values of the table on the document that is scanned for the batch and compare it with the parameter values.

- The first parameter specifies the heading line (COVERAGES) in the table to start.
- The second parameter specifies the row (BODILY INJURY LIABILITY).
- The third parameter specifies the column (LIMITS) to extract the cell data from the row and column intersections.
- Fourth and fifth parameters specify the field name for the row (Bodily Injury) and for the cell (Limits) to show on the document hierarchy.

If there are multiple rows with the similar data, multiple cells that meet the locate criteria are identified. In this scenario, multiple dynamic fields are created and populated in the DCO (One field for each row and a child field for each cell).

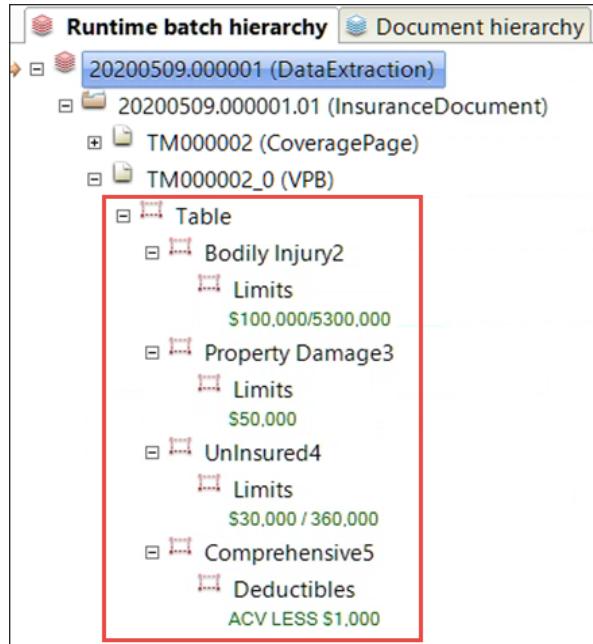
- The ValueStoreLocation parameter specifies the name of the DCO field (Table) that contains the extracted data.

- The last parameter controls the return value of the action if the search fails to find a result. When you set the value to False, a failure to find a search result still causes the action to return true. The return helps to ignore missing search results, allow the application to do multiple searches in a single function, and continue with following actions even if a value is not found.

You call this action several times for each row.



Test the table extraction



Creating a virtual page block and extracting values from tables

© Copyright IBM Corporation 2020

Figure 11-10. Test the table extraction

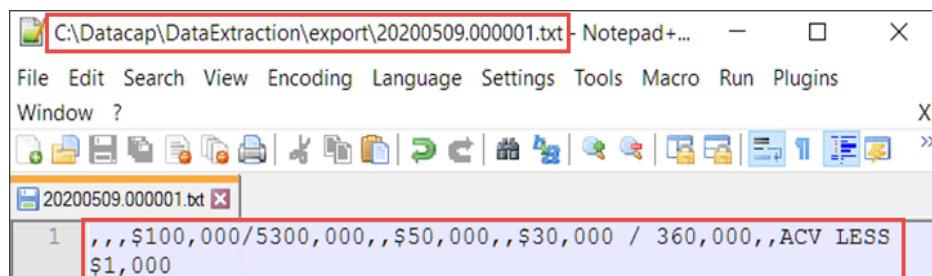
After you configure the rulesets for table extraction, test the application in the Datacap Studio Test tab.

In the Runtime batch hierarchy tab, the Table field in the VPB is populated with the extracted values. Subfields are created for each row in the table. In the screen capture, the values are shown in green text.

IBM Training 

Export the table data

COVERAGES	LIMITS	DEDUCTIBLES
BODILY INJURY LIABILITY	\$100,000 / \$300,000	
PROPERTY DAMAGE LIABILITY	\$50,000	
UNINSURED & UNDERINSURED MOTORISTS	\$30,000 / 360,000	
COMPREHENSIVE		ACV LESS \$1,000



Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-11. Export the table data

You can export the data that you extract from scanned documents by using Datacap.

The upper screen capture shows the table on the scanned document. The values that are extracted by Datacap are highlighted.

The lower screen capture shows the export document with the extracted values. This scenario shows the example of data export to a text file.

Unit summary

- Describe how to create a virtual page block
- Learn how to configure data extraction from tables

Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-12. Unit summary

Review questions

1. You set the value of True or False to the last parameter (FailMissingResults) of the FindTableValueRegEx action. Which of the following statements best describes the outcome when you set it to False?
 - A. If the search conditions do not find a match, Datacap ignores the missing search results, allows the application to run multiple searches in a single function, and continues with the next action
 - B. If there is an error other than a search not found, such as the page layout not containing a table, Datacap ignores the errors and allows the application to continue
 - C. If the search conditions do not find a match, Datacap stops the application to run multiple searches in a single function and stops the next action
 - D. If there is an error in creating a virtual page block, Datacap ignores the errors, allows the application to run multiple searches in a single function, and continues with the next action.



Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-13. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. You set the value of True or False to the last parameter (FailMissingResults) of the FindTableValueRegEx action. Which of the following statements best describes the outcome when you set it to False?
 - A. If the search conditions do not find a match, Datacap ignores the missing search results, allows the application to run multiple searches in a single function, and continues with the next action
 - B. If there is an error other than a search not found, such as the page layout not containing a table, Datacap ignores the errors and allows the application to continue
 - C. If the search conditions do not find a match, Datacap stops the application to run multiple searches in a single function and stops the next action
 - D. If there is an error in creating a virtual page block, Datacap ignores the errors, allows the application to run multiple searches in a single function, and continues with the next action



The answer is A.

Figure 11-14. Review answers

Exercise: Creating a virtual page block and extracting values from tables

Creating a virtual page block and extracting values from tables

© Copyright IBM Corporation 2020

Figure 11-15. Exercise: Creating a virtual page block and extracting values from tables

Exercise objectives

- Create a virtual page block
- Extract data from a table
- Test the application and verify the extracted data from the table



Creating a virtual page block and extracting values from
tables

© Copyright IBM Corporation 2020

Figure 11-16. Exercise objectives

Unit 12. Extracting data from label-value pairs

Estimated time

00:30

Overview

This unit explains how to extract values from label-value pairs.

How you will check your progress

- Review
- Complete the exercise

References

https://www.ibm.com/support/knowledgecenter/en/SSZRWV_9.1.7/com.ibm.datacapoc.doc/datacap_9.1.7.htm

Unit objectives

- Learn how to configure the Datacap application for data extraction from label-value pairs

Figure 12-1. Unit objectives

Topics

- Configure the Datacap application for data extraction from label-value pairs

Figure 12-2. Topics

12.1. Configure the Datacap application for data extraction from label-value pairs

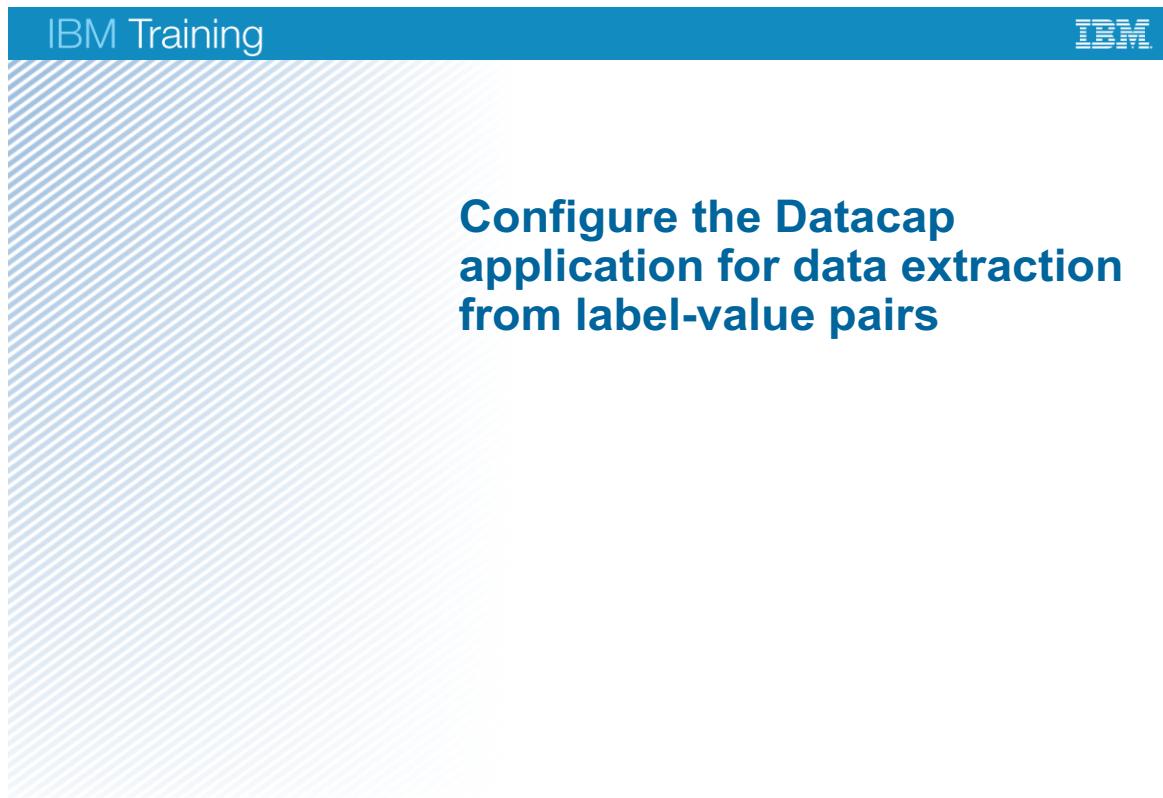


Figure 12-3. Configure the Datacap application for data extraction from label-value pairs

IBM Training 

Sample scan image for label-value pairs

GENERAL INSURANCE COMPANY
VERIFICATION OF COVERAGE
(SEE BELOW UNDER CAUTIONARY NOTE)

INSURED

ROBERT HANSON
1234 Oak St.
McKinney, CA, 75070

Policy Number: 902-319-444712
 Policy Date: From 04/08/2017 To 10/08/2017
 Registered State: TEXAS

To whom it may concern:
This letter is to verify that we have issued the policyholder coverage under the above policy number for the dates indicated in the effective and expiration date fields for the vehicle listed. This should serve as proof that the below mentioned vehicle meets or exceeds the financial responsibility requirement for your state.

This verification of coverage does not amend, extend or alter the coverage afforded by this policy.

Vehicle Year: 2015
 Make: BMW
 Model: 528i
 VIN: WBSFH4419F83241810

COVERAGES	LIMITS	DEDUCTIBLES
BODILY INJURY LIABILITY	\$100,000 / \$300,000	

Extracting data from label-value pairs

© Copyright IBM Corporation 2020

Figure 12-4. Sample scan image for label-value pairs

The screen capture shows a sample document that is used for scanning and the label-value pairs are highlighted.

- Upper part of the page: Policy Number, Policy Date, and Registered State
- Lower part of the page: Vehicle Year, Make, Model, and VIN

Find the label-value pairs from the layout XML file

- The ruleset contains actions for finding label value pairs

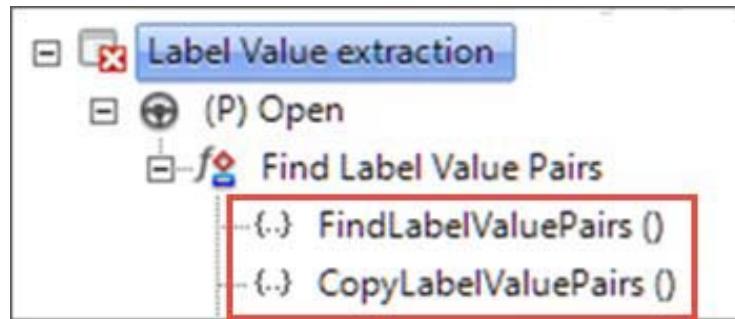


Figure 12-5. Find the label-value pairs from the layout XML file

The screen capture shows a ruleset that contains the following actions:

FindLabelValuePairs ()

- The action searches the layout XML file for labels and their associated values and returns the results. The data in the results can be used by other actions to populate fields.
- The results from the action are also saved to the layout XML file. FindLabelValuePairs requires a previously created layout file.
- The label and values pairs are determined by using font attributes and relative positioning of labels to values to identify what is a label and what is a value.

CopyLabelValuePairs ()

- Creates fields in the batch hierarchy for each label-value pair that is found by the FindLabelValuePairs ()
- Is helpful during the development to check that the FindLabelValuePairs () action returned values
- Requires a previously created layout file



Test the Find label value pairs configuration

Runtime batch hierarchy Document hierarchy

- 20200511.000000 (DataExtraction)
 - 20200511.000000.01 (InsuranceDocument)
 - TM000002 (CoveragePage)
 - Layout
 - PolicyNumber
 - State
 - VehicleYear
 - Make
 - Model
 - VIN
 - LabelValue-0
902-319-444712
 - LabelValue-1
From 04/08/2017 To 10/08/2017
 - LabelValue-2
TEXAS
 - LabelValue-3
2015
 - LabelValue-4

Policy Number: 902-319-444712
Policy Date: From 04/08/2017 To 10/08/2017
Registered State: TEXAS

Vehicle Year: 2015
Make: BMW
Model: 528i
VIN: WBSFH4419F83241810

Extracting data from label-value pairs © Copyright IBM Corporation 2020

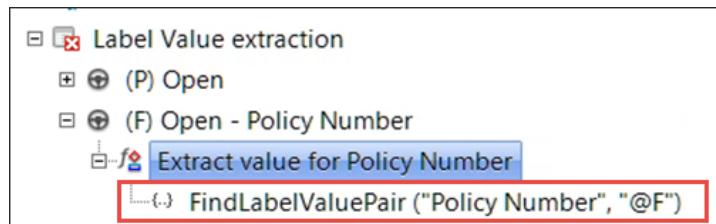
Figure 12-6. Test the Find label value pairs configuration

After you configure the rulesets for finding label-value pairs, test the application in Datacap Studio.

The left screen capture shows the Runtime batch hierarchy tab in the Datacap Studio Test tab. The Datacap fields are populated with the extracted data from label-value pairs. The data is shown in green text.

For comparison, the right screen captures show the label-value pairs on the scanned page.

Extract the data from a specific label-value pair



Extracting data from label-value pairs

© Copyright IBM Corporation 2020

Figure 12-7. Extract the data from a specific label-value pair

The screen capture shows a field level rule that contains the `FindLabelValuePair ()` action.

The action populates a field with a value that is found by the `FindLabelValuePairs` action.

The first parameter (Policy Number) identifies the label for which you want to retrieve the value. The second parameter identifies the target object to save the value. In this example, it uses a field.



Test the data extraction for specific label-value pairs

The screenshot displays the Datacap Studio interface. On the left, the 'Runtime batch hierarchy' tab is active, showing a tree structure of data. A red box highlights the 'PolicyNumber' field, which is populated with the value '902-319-444712'. To the right, two boxes show the extracted data: one for vehicle details and one for a VIN number.

Label	Value
Policy Number	902-319-444712
Policy Date	From 04/08/2017 To 10/08/2017
Registered State	TEXAS
Vehicle Year	2015
Make	BMW
Model	528i
VIN	WBSFH4419F83241810

Figure 12-8. Test the data extraction for specific label-value pairs

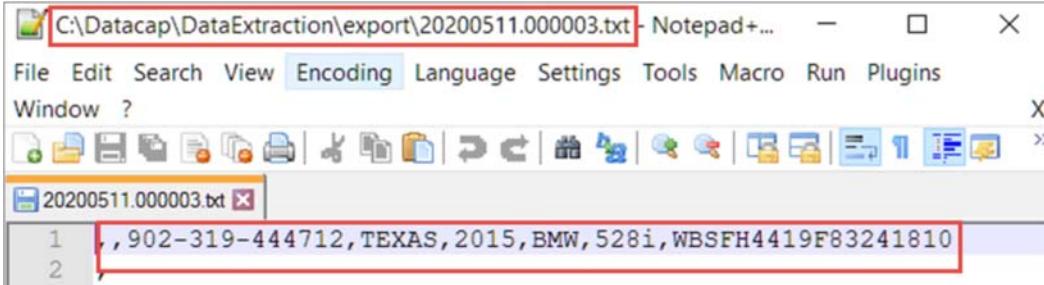
After you configure the rulesets, test the application in Datacap Studio.

The left screen capture shows the Runtime batch hierarchy tab in the Datacap Studio Test tab. The Datacap fields are populated with the extracted data from label-value pairs. The data is shown in green text. Notice that the Datacap fields have specific names as compared to the previous test that showed LabelValue-1.

For comparison, the right screen captures show the label-value pairs on the scanned page.

IBM Training 

Export the data from label-value pairs



The screenshot shows a Notepad+ window with the title bar "C:\Datacap\DataExtraction\export\20200511.000003.txt - Notepad+...". The menu bar includes File, Edit, Search, View, Encoding, Language, Settings, Tools, Macro, Run, Plugins, Window, and ?.

The main area displays two lines of data:

```
1 , 902-319-444712, TEXAS, 2015, BMW, 528i, WBSFH4419F83241810
2 ,
```

Below the Notepad+ window, there are two boxes containing extracted data:

Policy Number: 902-319-444712 Policy Date: From 04/08/2017 To 10/08/2017 Registered State: TEXAS	Vehicle Year: 2015 Make: BMW Model: 528i VIN: WBSFH4419F83241810
--	---

Extracting data from label-value pairs

© Copyright IBM Corporation 2020

Figure 12-9. Export the data from label-value pairs

The upper screen capture shows the export text file with the extracted data from label-value pairs. The lower screen capture shows the label-value pairs on the scanned document for comparison.

Unit summary

- Learn how to configure the Datacap application for data extraction from label-value pairs

[Extracting data from label-value pairs](#)

© Copyright IBM Corporation 2020

Figure 12-10. Unit summary

Review questions

1. Which of the following statements are TRUE about the FindLabelValuePairs action? (Choose two)
 - A. The label and value pairs are determined by using relative positioning of labels to values
 - B. The results from this action are saved to a previously created layout XML file
 - C. The action creates a layout XML file
 - D. The label and value pairs are determined by using keywords
 - E. The label and value pairs are determined by using Content Classification.
2. True or False: The CopyLabelValuePairs action creates fields in the batch hierarchy for each label value pair that was found.



Extracting data from label-value pairs

© Copyright IBM Corporation 2020

Figure 12-11. Review questions

Write your answers here:

- 1.
- 2.
- 3.

Review answers

1. Which of the following statements are TRUE about the FindLabelValuePairs action? (Choose two)
 - A. [CORRECT] The label and value pairs are determined by using relative positioning of labels to values
 - B. [CORRECT] The results from this action are saved to a previously created layout XML file
 - C. The action creates a layout XML file
 - D. The label and value pairs are determined by using keywords
 - E. The label and value pairs are determined by using Content Classification

The answer is A and B.
2. True or False: The CopyLabelValuePairs action creates fields in the batch hierarchy for each label value pair that was found.
The answer is True.



Figure 12-12. Review answers

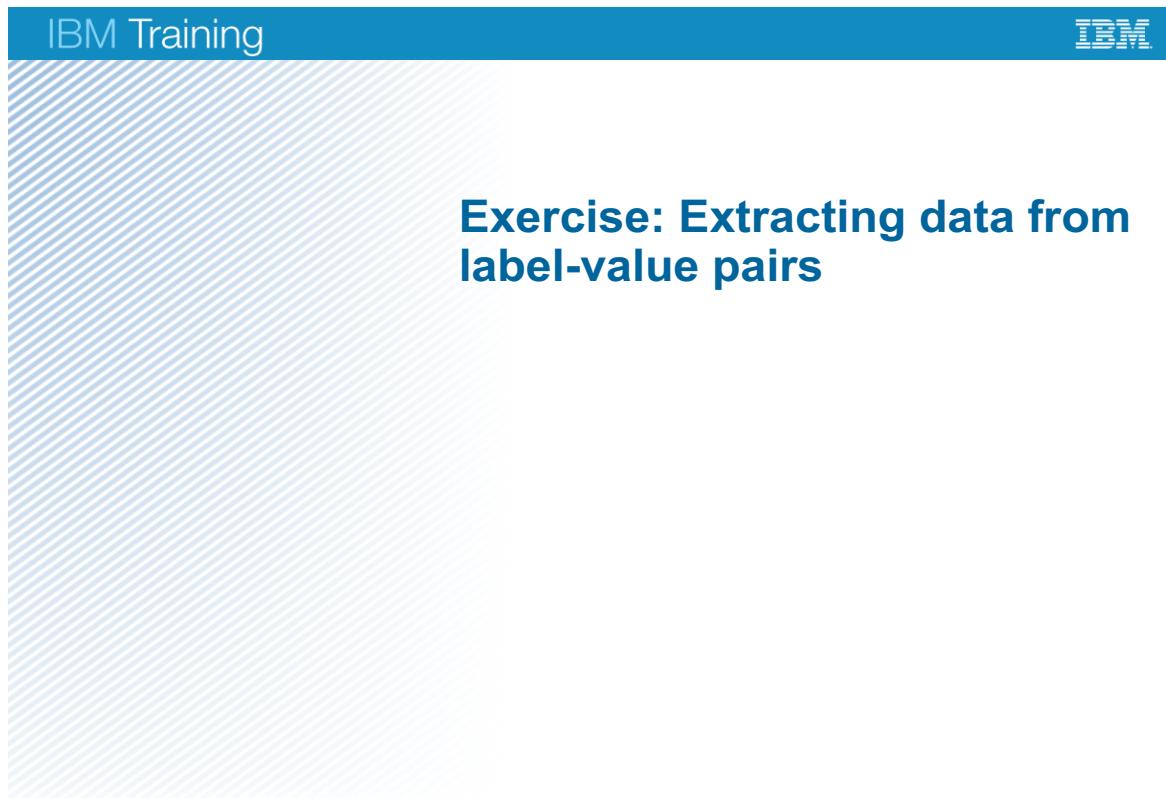


Figure 12-13. Exercise: Extracting data from label-value pairs

Exercise objectives

- Extract data from label-value pairs
- Test the application and verify the extracted data



Figure 12-14. Exercise objectives

Unit 13. Course summary, badge, and other learning resources

Estimated time

00:30

Overview

This unit summarizes the course and provides badges and other information for future study.

How you will check your progress

- Review

Unit objectives

- Describe the course objectives and what you learned
- Earn a badge for this course
- Identify and describe product certifications that are related to this course
- Identify resources that can help you learn more

© Copyright IBM Corporation 2020

Figure 13-1. Unit objectives

Course objectives (1 of 2)

- Describe capture concepts and Datacap
- Identify the tasks in Datacap process
- Describe the Datacap recognition methods that are used for scanned images
- Understand Datacap architecture
- Process batches in the Datacap clients
- Identify the application development features in Datacap Studio
- Describe the Datacap application design concepts
- Build a Datacap application with Forms Template
- Configure rulesets for a Datacap application
- Examine the logs files for debugging
- Troubleshoot the applications in the Datacap Studio Test tab

© Copyright IBM Corporation 2020

Figure 13-2. Course objectives (1 of 2)

Course objectives (2 of 2)

- Build an application for processing multiple page types in a batch
- Configure an application for recognition with OMR
- Export data and scanned documents to a FileNet Content Manager repository
- Configure a Datacap application to create page layouts
- Create a virtual page block
- Extract data from tables and label-value pairs

© Copyright IBM Corporation 2020

Figure 13-3. Course objectives (2 of 2)



IBM badge

- Earn a Skills badge for this course by passing a quiz
- To earn the badge for this course:
https://ibm-learning-skills-dev.github.io/ibm-learning-skills-dev.github.io/badges/IBM_Datacap_V9.1.4-V9.1.7.html
- Other IBM Cloud badges:
<https://ibm-learning-skills-dev.github.io/ibm-learning-skills-dev.github.io/badges/badgeindex.html>

© Copyright IBM Corporation 2020

Figure 13-4. IBM badge

IBM Professional Certifications

- By achieving an IBM Professional Certification, you can demonstrate your IBM Cloud product mastery to your employer or clients
- Certifications are a higher level of credential than a Skills badge for a single education course
- Product certifications demonstrate a strong knowledge of the product and typically require several months of work with the product
- IBM Cloud certifications are available for several roles, including developers, administrators, and business analysts
- For information on specific certifications and their requirements, see <http://www.ibm.com/certify>

© Copyright IBM Corporation 2020

Figure 13-5. IBM Professional Certifications

Other learning resources (1 of 4)

- **IBM Skills Gateway**

- Search the new IBM Training and Skills website (formerly IBM Authorized Training website) to find and access the content that you want.
- <https://www.ibm.com/training>

- **IBM Cloud Education Course Information Home**

- Go to the wiki to find course abstracts, course correction documents, and curriculum development plans for IBM Cloud offerings.
- <https://ibm-learning-skills-dev.github.io/ibm-learning-skills-dev.github.io/education/courseinfo.html>

- **Role-based Learning Journeys**

- Learning Journeys describe the appropriate courses, in the recommended order, for specific products and roles.
- https://www.ibm.com/services/learning/journey_category?categoryId=o-itns-01-02

© Copyright IBM Corporation 2020

Figure 13-6. Other learning resources (1 of 4)

Other learning resources (2 of 4)

- **IBM Professional Certification Program**

- IBM Professional Certification enables skilled IT professionals to demonstrate their expertise to the world. It validates skills and proficiency in the latest IBM technology and solutions.
- <https://www.ibm.com/certify>

- **IBM Training blog, Twitter, and Facebook**

- These official IBM Training and Skills accounts provide information about IBM course offerings, industry information, conference events, and other education-related topics.
- <https://www.ibm.com/blogs/ibm-training>
- <https://twitter.com/IBMTTraining>
- <https://www.facebook.com/ibmtraining>

© Copyright IBM Corporation 2020

Figure 13-7. Other learning resources (2 of 4)

Other learning resources (3 of 4)

- **Business Partner Technical Enablement Portal**
 - <https://ibm.box.com/s/695khv9nyzekaorykqmsjrematz3v9xh>
 - This program provides technical training content modules to IBM software partners (through PartnerWorld) and IBM Business Partners.
- **IBM Developer**
 - IBM's official developer program offers access to software trials and downloads, how-to information, and expert practitioners.
 - <https://developer.ibm.com/>
 - <https://github.com/ibm-eclm/datacap-developer-kit>
- **IBM Education Assistant**
 - These multimedia educational modules help users gain a better understanding of IBM Software products and use them more effectively to meet business requirements.
 - <https://www.ibm.com/products/software>

© Copyright IBM Corporation 2020

Figure 13-8. Other learning resources (3 of 4)

Other learning resources (4 of 4)

- **IBM Knowledge Center**

- The IBM Knowledge Center is the primary home for IBM product documentation.
- <https://www.ibm.com/support/knowledgecenter>

- **IBM Marketplace**

- IBM Marketplace is the landing page for all IBM Cloud products. Go to the Marketplace to learn about IBM offerings for Cloud, Cognitive, Data and Analytics, Mobile, Security, IT Infrastructure, and Enterprise and Business Solutions.
- <https://www.ibm.com/products>

- **IBM Redbooks**

- IBM Redbooks are developed and published by the IBM International Technical Support Organization (ITSO). Redbooks typically provide positioning and value guidance, installation and implementation experiences, typical solution scenarios, and step-by-step "how-to" guidelines.
- <http://www.redbooks.ibm.com/>

© Copyright IBM Corporation 2020

Figure 13-9. Other learning resources (4 of 4)

Unit summary

- Describe the course objectives and what you learned
- Earn a badge for this course
- Identify and describe product certifications that are related to this course
- Identify resources that can help you learn more

© Copyright IBM Corporation 2020

Figure 13-10. Unit summary

IBM Training

IBM

Course completion

You have completed this course:
Developing Applications in IBM Datacap V9.1.7



Do you have any questions?

© Copyright IBM Corporation 2020

Figure 13-11. Course completion



IBM Training



© Copyright International Business Machines Corporation 2020.