WASHINGTON, DC | JUNE 26–27, 2024

aws SUMMIT

Building an intelligent document processing hub at scale at Fannie Mae

Noah Donaldson

Sr. Solutions Architect

AWS

Daniel Ricardo Holstein

VP, Technology and Shared Services

Fannie Mae

Our mission

To support the customer's mission with solution guidance, architectures, best practices, highly reusable POCs, and solution starters. We work backwards from customer's outcomes and build technical assets to support their mission.

Document workflow automation

Onboard new document types through low-code no-code (LCNC)

Advanced extraction techniques using Amazon Textract
Event-driven workflow to trigger downstream process

Customizable and extensible

Intuitive and easy-to-use interface for business users
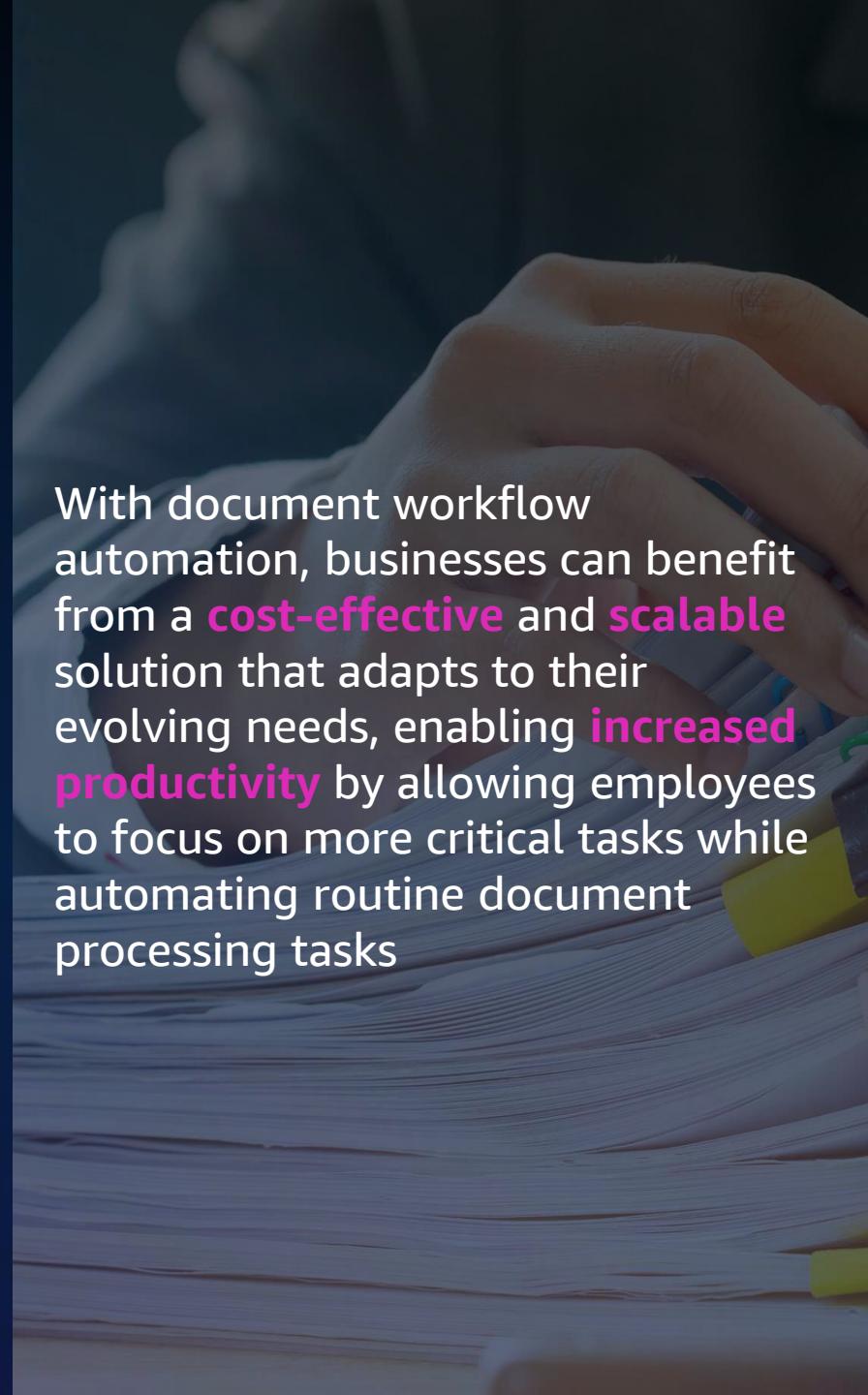
With document workflow automation, businesses can benefit from a **cost-effective** and **scalable** solution that adapts to their evolving needs, enabling **increased productivity** by allowing employees to focus on more critical tasks while automating routine document processing tasks

Our approach
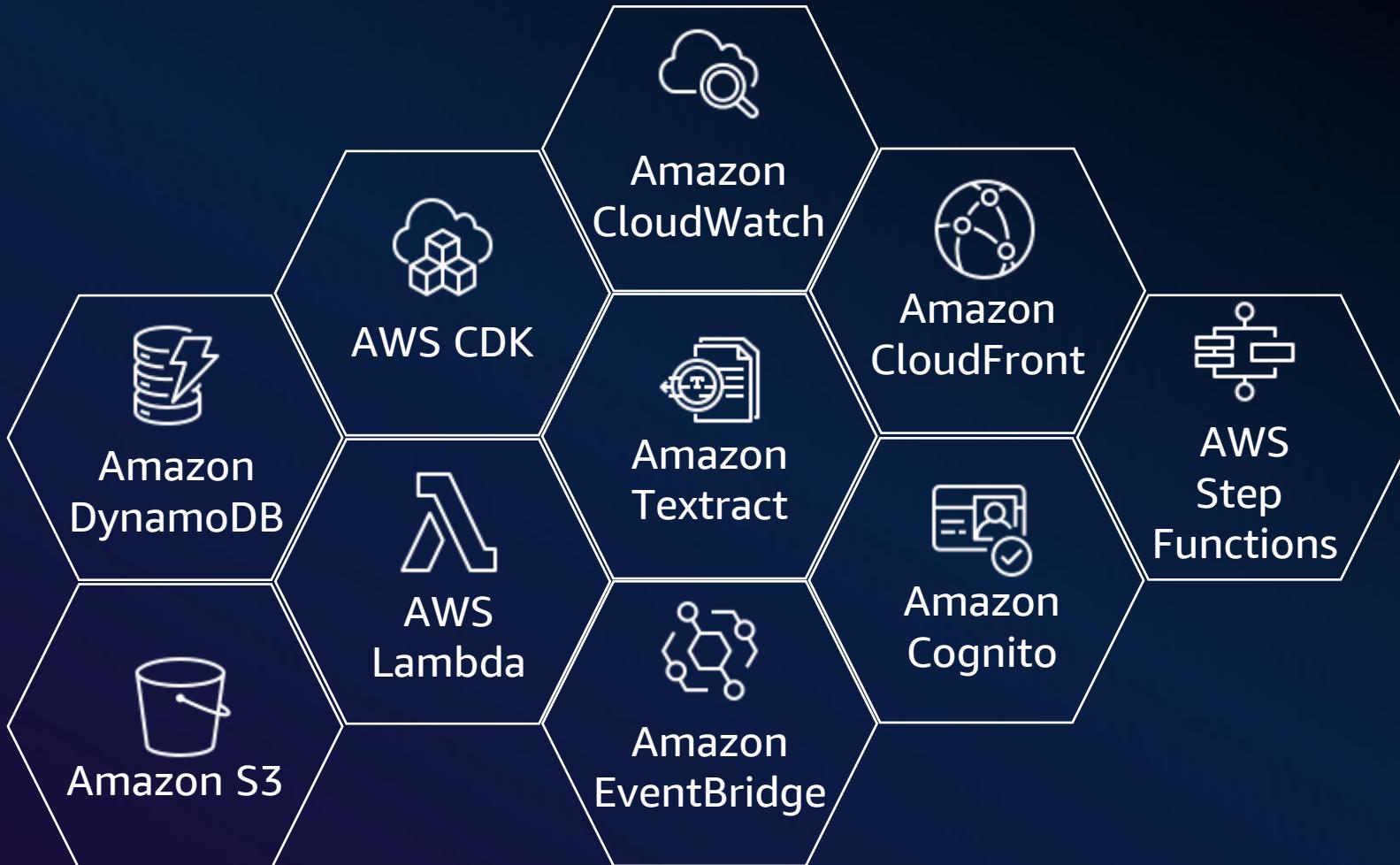
Amazon DynamoDB

Amazon CloudWatch

Amazon CloudFront

Amazon EventBridge

AWS CDK

Amazon Cognito

Amazon S3

AWS Lambda

Amazon Textract

AWS Step Functions

# Our approach

Document workflow automation
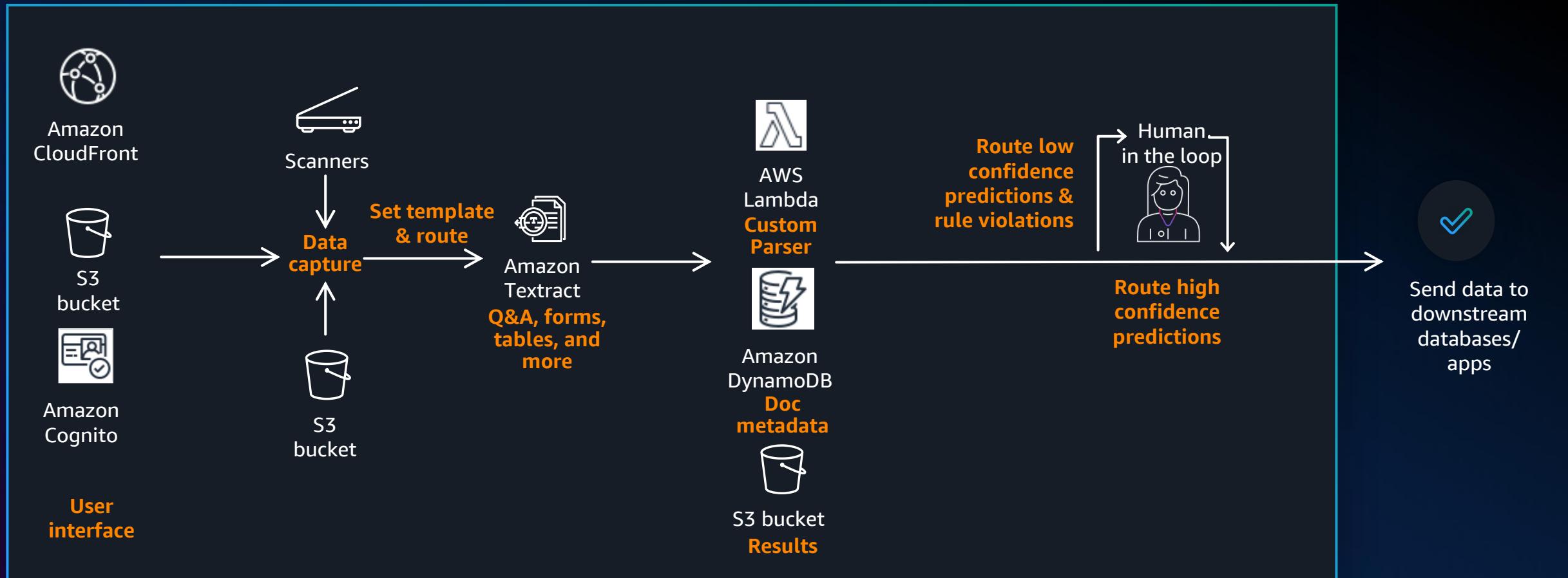
Amazon CloudFront

S3 bucket

Amazon Cognito

User interface

Scanners

Data capture

Set template & route

S3 bucket

Amazon Textract

Q&A, forms, tables, and more

AWS Lambda
Custom Parser

Amazon DynamoDB
Doc metadata

S3 bucket
Results

Route low confidence predictions & rule violations

Human in the loop

Route high confidence predictions

Send data to downstream databases/ apps

Onboard new documents, manage configurations, & verify documents

Receive documents

Extract data from documents

Parse & store data

Verification & human review

Derive key insights with extraction & enrichment

Document workflow automation

Amazon CloudFront

S3 bucket

Amazon Cognito

User interface

Scanners

Data capture

Set template & route

S3 bucket

Amazon Textract

Q&A, forms, tables, and more

AWS Lambda **Custom Parser**

Amazon DynamoDB **Doc metadata**

S3 bucket **Results**

Route low confidence predictions & rule violations

Human in the loop

Route high confidence predictions

Send data to downstream databases/ apps

Onboard new documents, manage configurations, & verify documents

Receive documents

Extract data from documents

Parse & store data

Verification & human review

Derive key insights with extraction & enrichment

DWA – Onboardings & rules

1.

Upload Document

- Upload Document
- Configure Settings
- Review Fields
- Configure Queries

The onboarding tool allows you to introduce new document types to our system. For each specific document type, such as the 1099 or 1040 tax forms, the tool configures the system to recognize its structure and content.

Upload a file for onboarding.

Choose file

(PNG, JPEG, TIFF, or PDF)

0ddff947-504b-45f3-925b-9434ccd749f7.jpeg
143.13 KB
2024-03-28T09:46:42

Name this document type.

W-2

Next

2.

Configure Settings
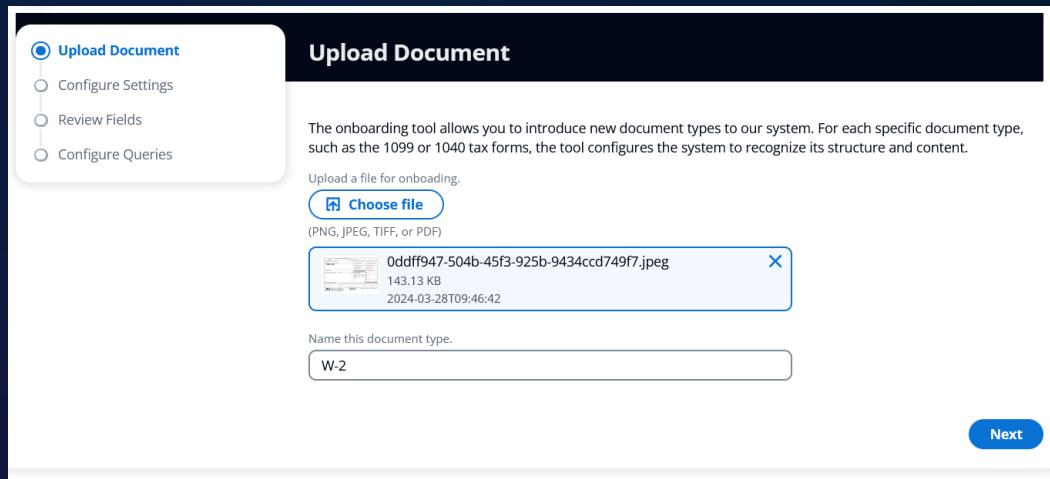
- Upload Document
- Configure Settings
- Review Fields
- Configure Queries

Pre-Extraction Process
Process to run prior to sending document to Textract. See Onboarding Documentation for more information.

Default

Form Confidence Flag Level
If any Key/Value, Query, or Table Row is below this value, this document will be flagged.

50

Extraction Types
Type of extraction to perform. Forms is Key-Value extraction. Queries allows the user to write custom natural language queries. Finally, Tables extracts data from tables in the document.

- ☑ Forms
- ☐ Tables
- ☑ Queries

Data Parser
A process to parse data after data has been extracted from the document.

Default

Previous   Next

3.

Review Fields

- Upload Document
- Configure Settings
- Review Fields
- Configure Queries
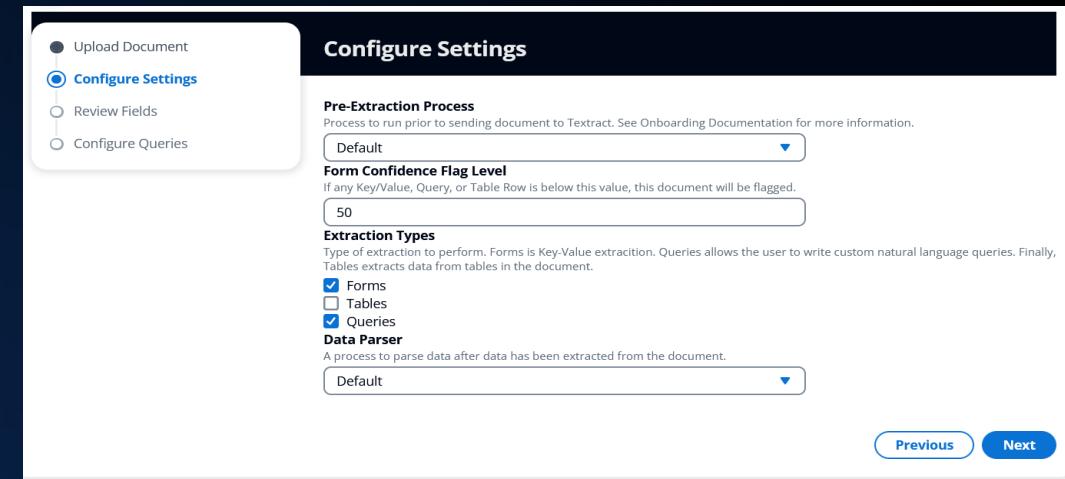
Filter...   < 1 > ⚙

Field Name	Extracted Value
12a	
Suff.	
f Employee's address and ZIP code	
12b	
5 Medicare wages and tips	
7 Social security tips	
16 State wages, tips, etc.	
Employer's state ID number	
11 Nonqualified plans	

4.

Configure Queries
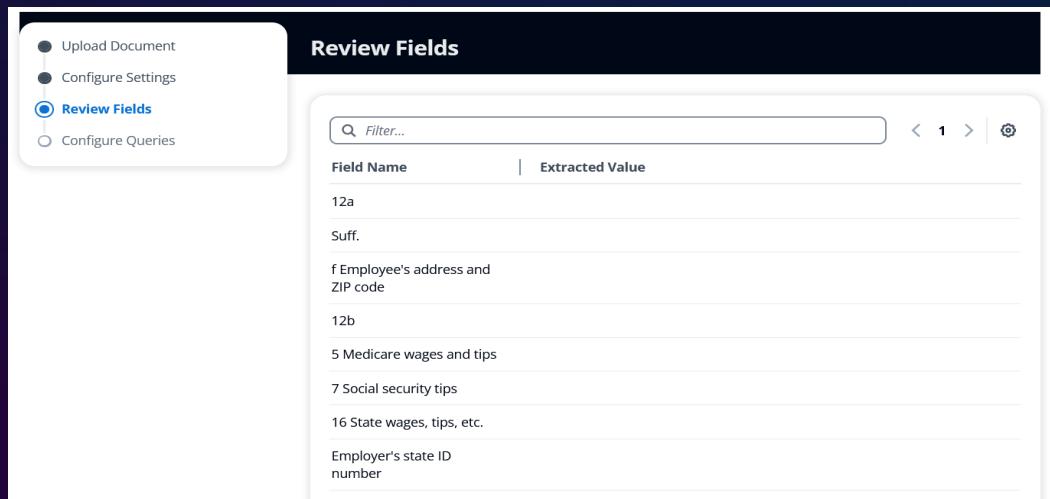
- Upload Document
- Configure Settings
- Review Fields
- Configure Queries

0ddff947-504b-45f3-925b-9434ccd749f7.jpeg

📍 Test Queries   Add   Delete

Filter...   < 1 > ⚙

	Query	Pages	Result
○	What is the OMB Number	1	1545-0008

Previous   Submit

Document workflow automation

Amazon CloudFront

S3 bucket

Amazon Cognito

User interface

Scanners

Data capture

S3 bucket

Set template & route

Amazon Textract

Q&A, forms, tables, and more

AWS Lambda **Custom Parser**

Amazon DynamoDB **Doc metadata**

S3 bucket **Results**

Route low confidence predictions & rule violations

Human in the loop

Route high confidence predictions

Send data to downstream databases/ apps

Onboard new documents, manage configurations, & verify documents

Receive documents

Extract data from documents

Parse & store data

Verification & human review

Derive key insights with extraction & enrichment

DWA – Human review

Document workflow automation

Ingest	ML-based ingest – can be trained to automatically classify sections of multi-page docs
Onboarding	Onboard new document types through LCNC UI
Extraction	Advanced extraction types (forms, tables, & queries) Customizable extraction process
Validation	Confidence evaluation of extracted data Customizable data validation
Human review	Interact with extracted documents Highlight fields on a document Summarize violations for quick correction
Downstream business process	Event-driven workflow to trigger downstream process

Fannie Mae

86th year
supporting
homeowners & renters

$4 trillion
financial institution

$369 billion
in liquidity
provided in 2023

Fannie Mae is a leading provider of mortgage financing in the US

We purchase mortgages from lenders to free up the money they need to make other mortgage loans, therefore ensuring the ongoing availability of affordable mortgages

# Fannie Mae processes large volumes of documents every year

By the numbers

Q1 of 2024

155,000
single-family purchase loans acquired

36,000
single-family refinance loans acquired

89,000
units of multifamily rental housing acquired

Need for automated extraction of data

📄 **Loan files can be up to 1,000 pages each; examples include HUD1, insurance, appraisal, credit report, payment verification, and more**

🔍 **Manual extraction is a multistep process that is time-consuming and prone to errors**

⚙️ **Automation saves thousands of hours of manual extraction and allows for interfacing with upstream and downstream systems**

Ingest documents

Capture

Verify

Classify

Validate

Export
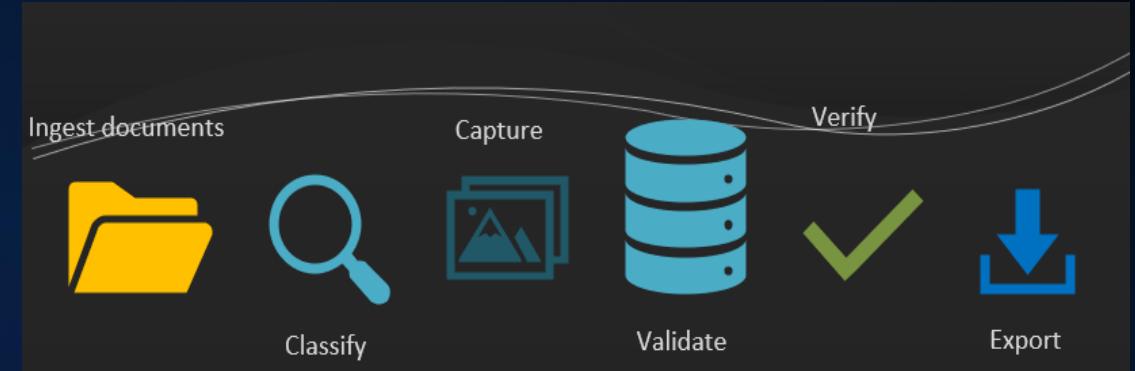
# Challenges with extracting information

Scenarios

- ✓ Multiple document types within the same batch

- ✓ Format variation within document types

- ✓ Handwritten, unstructured documents

- ✓ Multiple document types within a single document/page

- ✓ Poor-quality documents (faxed, phone images)

- ✓ Identification and extraction of multi-line expenses

- ✓ Old manuscripts
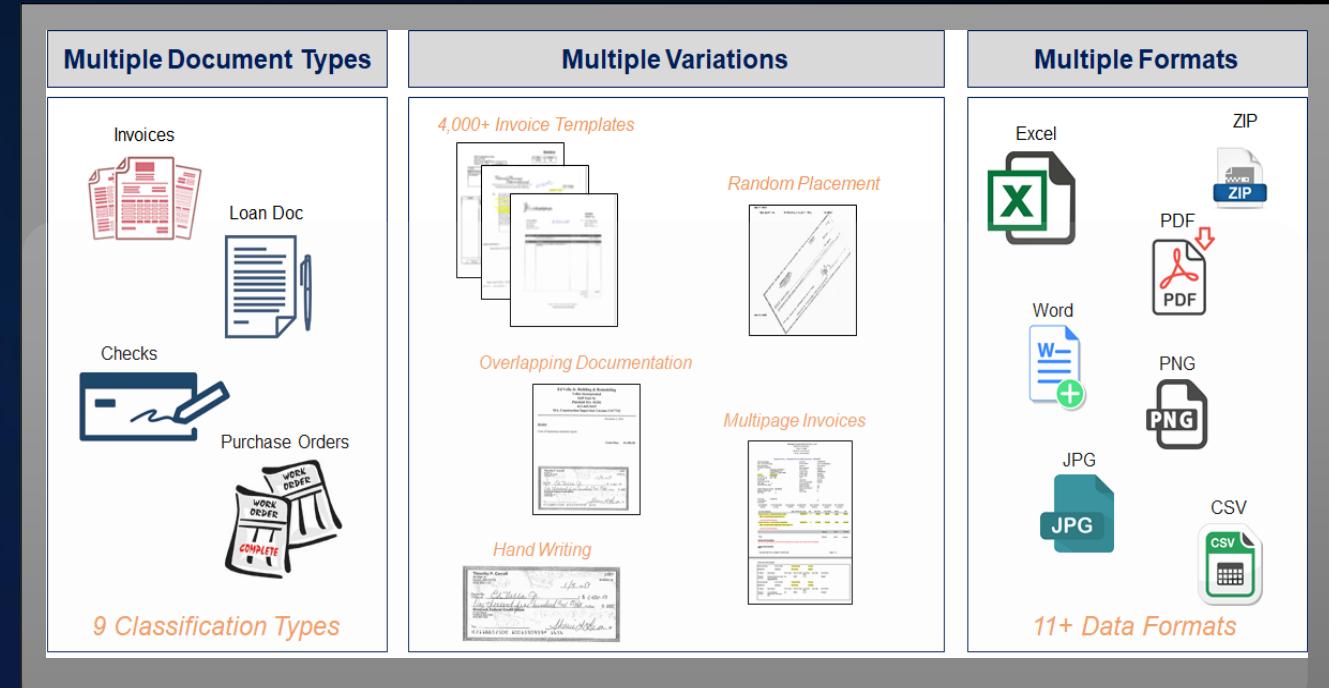
Our collaboration with AWS

Fannie Mae is working with AWS to enhance their classification, extraction, and orchestration capabilities

- Users can quickly onboard new document types and obtain extraction results with self-service option
- We are developing cutting-edge extraction capabilities capable of processing highly unstructured data with option for human review based on confidence levels
- Fully resilient, serverless architecture

Upload documents via portal

Classify/Extract

Human In the Loop

Manual review (if needed)

Extracted Data
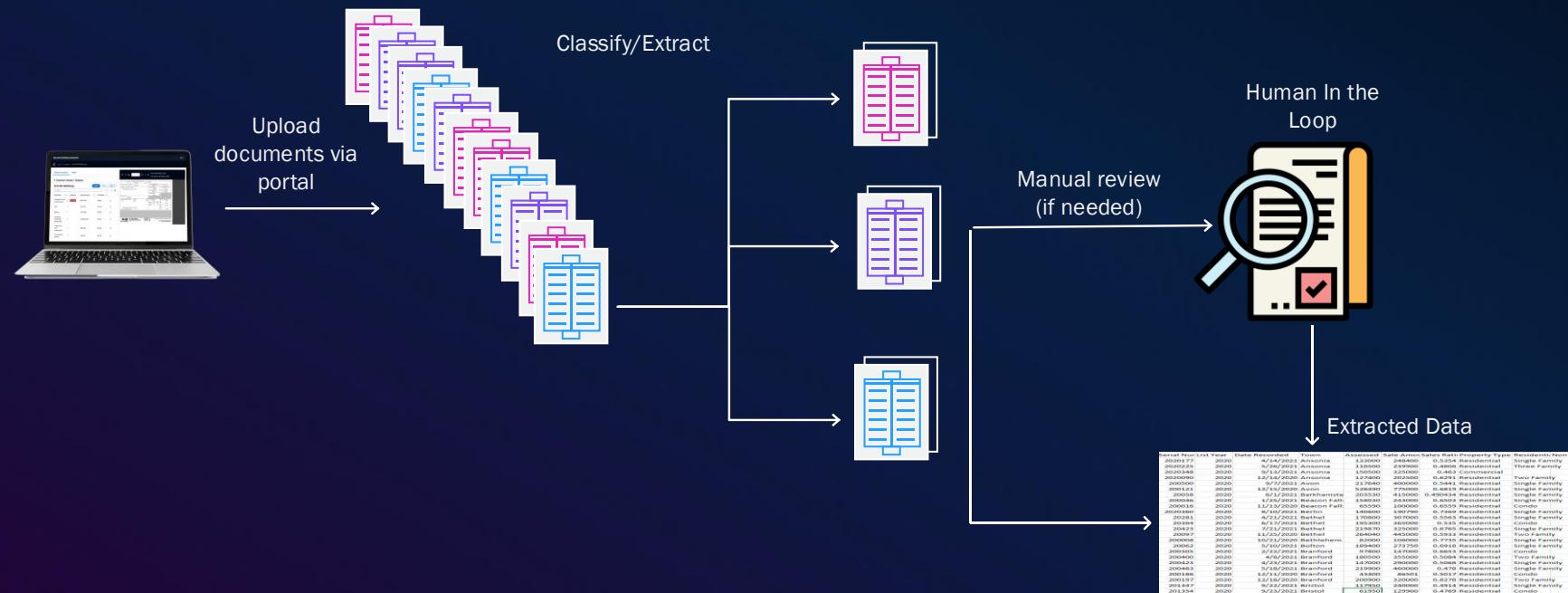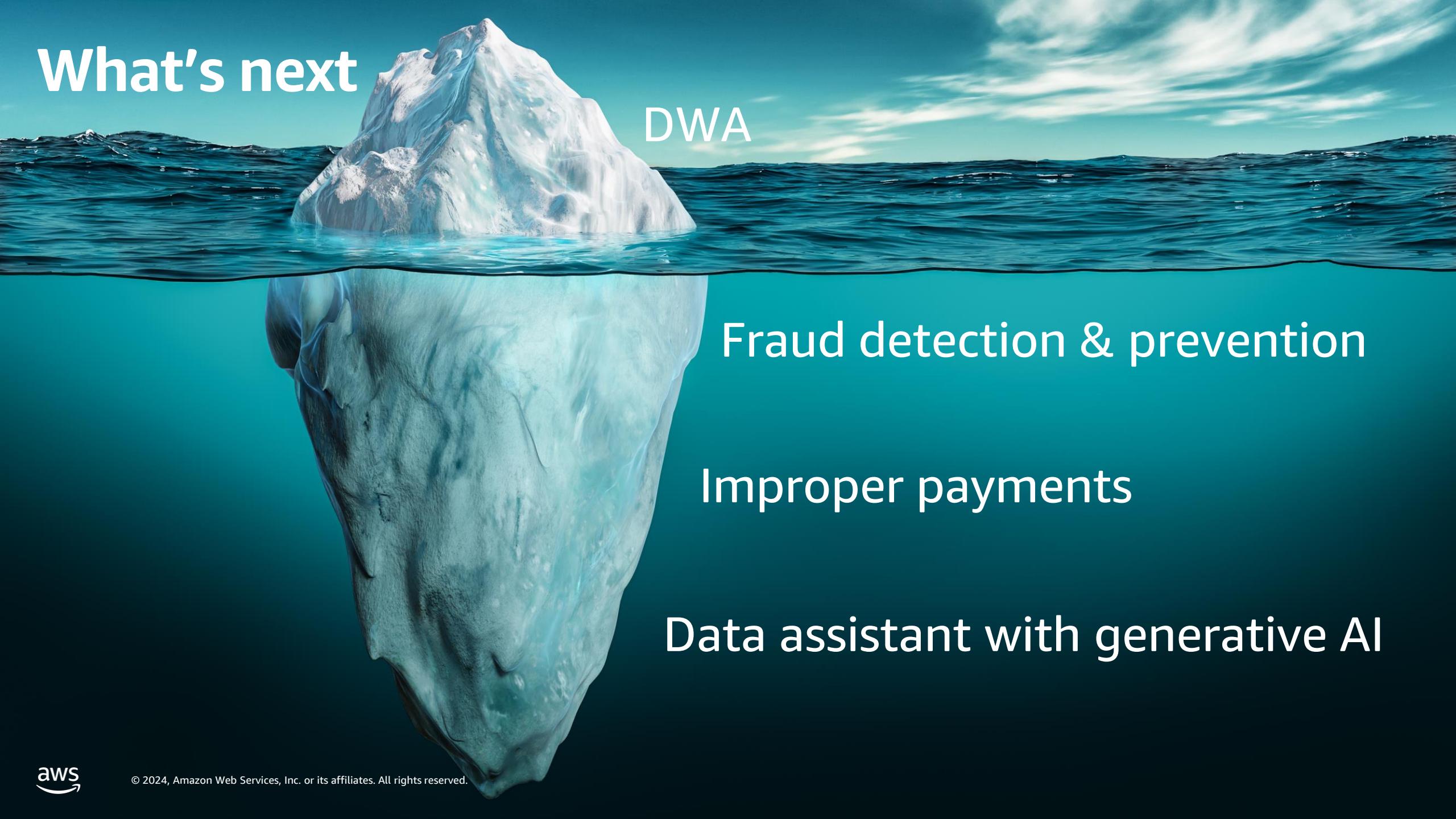
What's next

DWA

Fraud detection & prevention

Improper payments

Data assistant with generative AI

skillbuilder.aws

Build beyond
Create a free account
on AWS Skill Builder to
gain in-demand skills