

AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

AIM202

Automate your mortgage document processing with AWS AI/ML

Jane Han

Sr. Product Manager Tech
Amazon Web Services

Eric Barrow

SVP, Application Development
Pennymac



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Today's challenges in legacy document processing

What is intelligent document processing (IDP)?

Amazon Textract overview

What we are launching

Why **Amazon Textract Analyze Lending?**

- Mortgage industry overview
- Today's challenges in the mortgage industry

Pennymac case study

- How we solve the problem

IDP AWS Partners

Important resources and links

Legacy document processes do not meet today's needs



Legacy OCR and manual processes are time-consuming, error-prone, and expensive



Manual processes do not scale easily with document volume



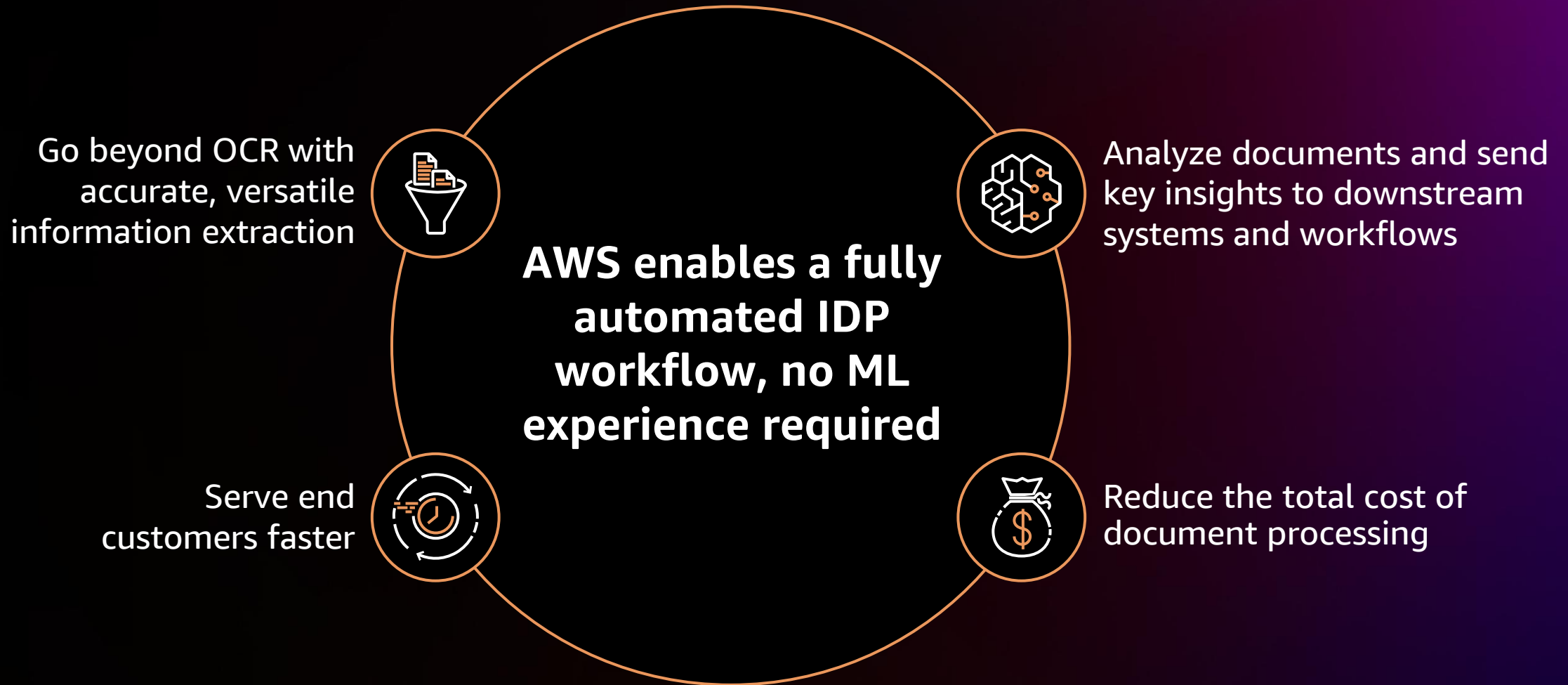
Difficult to find useful information needed for business decisions

Faster document processing shortens decision cycles;
this allows you to serve more customers and have people do higher-value tasks

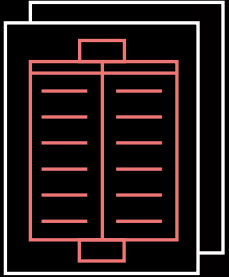
The hidden costs of legacy systems



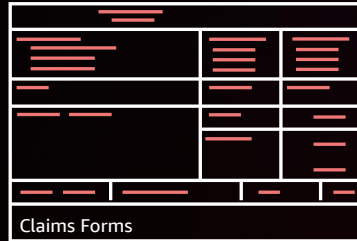
Benefits of AWS intelligent document processing



Amazon Textract – Features

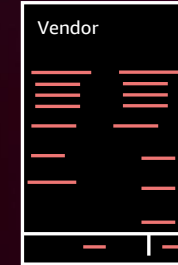


Text (OCR)

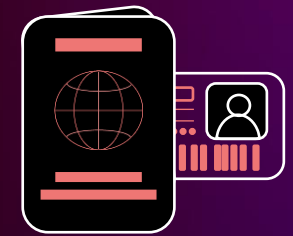


Forms

Invoices &
receipts



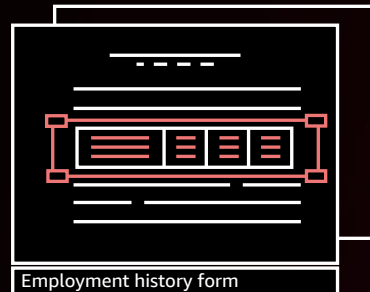
Identity documents



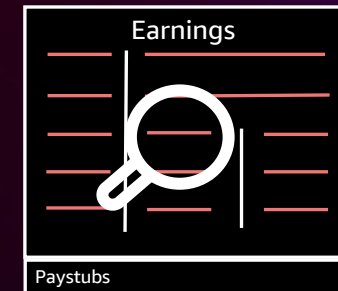
Specialized documents



Handwriting



Tables



Queries

Go beyond OCR with accurate, versatile information extraction

Accuracy of forms and tables allows customers to automate documents that previously needed manual processing

INPUT → OUTPUT

I. TYPE OF MORTGAGE AND TERMS OF LOAN				
Mortgage Applied for: <input type="checkbox"/> VA <input checked="" type="checkbox"/> Conventional <input type="checkbox"/> Other (explain):		Agency Case Number		Lender Case Number
<input type="checkbox"/> FHA <input type="checkbox"/> USDA/Rural Housing Service		ABC1234		XYZ6543
Amount	Interest Rate	No. of Months	Amortization Type: <input checked="" type="checkbox"/> Fixed Rate <input type="checkbox"/> Other (explain):	
\$ 552,500	3.5 %	360	<input type="checkbox"/> GPM <input type="checkbox"/> ARM (type):	
II. PROPERTY INFORMATION AND PURPOSE OF LOAN				
Subject Property Address (street, city, state, & ZIP)				No. of Units
123 Any Street, Anytown, USA, 12345				1
Legal Description of Subject Property (attach description if necessary)				Year Built
Single Family Home				2015

Gross Monthly Income	Borrower	Co-Borrower	Total
Base Empl. Income*	\$ 85,000	\$ N/A	\$ 85,000
Overtime	10,000	N/A	10,000

Raw text

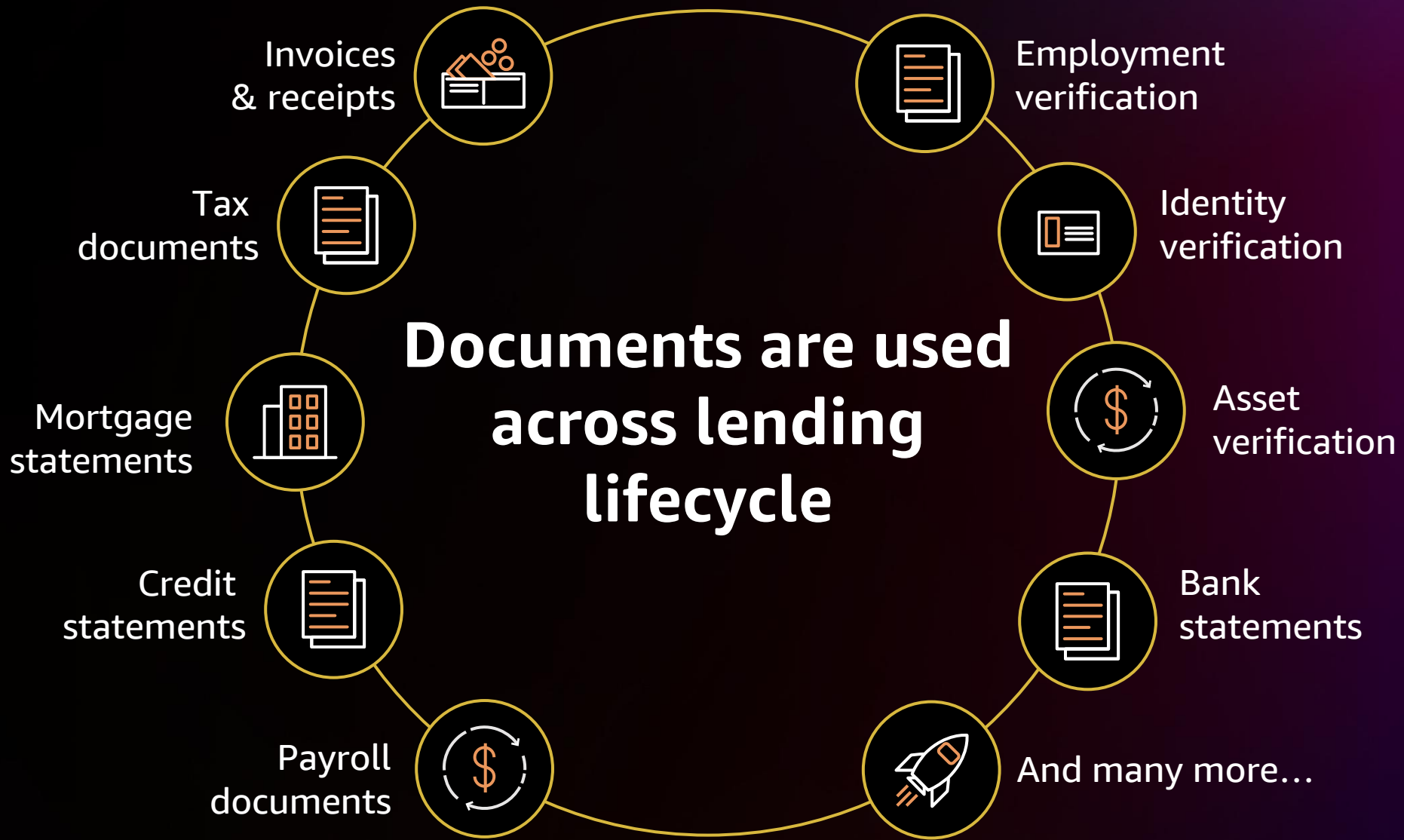
Borrower Co-Borrower I. TYPE OF MORTGAGE AND TERMS OF LOAN Mortgage VA Conventional
Other (explain): Agency Case Number Lender Case Number Applied for: FHA USDA/Rural
ABC1234 XYZ6543 Housing Service Amount Interest Rate No. of Months Amortization Type:
Fixed Rate Other (explain): \$ 552,500 3.5 % 360 GPM ARM (type):

Forms

Lender Case Number	VA
XYZ6543	NOT_SELECTED
Conventional	Other (explain):
SELECTED	

Table recognition

Column 1	Column 2	Column 3	Column 4
Gross Monthly Income	Borrower	Co-Borrower	Total
Base Empl. Income*	\$ 85,000	\$ N/A	\$ 85,000
Overtime	10,000	N/A	10,000



Challenges in mortgage loan data extraction

Document classification

(mortgage statement vs. payroll document vs. identity document)

- Pre processing
- Expensive & time consuming
- Custom models

Data variations

(date of birth vs. DOB vs. birthday)

- Post processing
- Expensive & time consuming

Data structure

(table, form, implied fields)

- Post processing
- Custom models

Nested data

(page sections, duplicate fields)

- Development & management overhead
- Templates are brittle

Routing & orchestration

(choosing the best data extraction model for the document type)

- Development & management overhead
- Expensive & time consuming

Packaged use-case solution to accelerate workflows

AMAZON Textract DOCUMENT PROCESSING SOLUTION FOR MORTGAGES

The image shows a stack of three documents. The top document is an 'Earnings Statement' with fields for CO, FILE, DEPT, CLOCK, NUMBER, and AGENCY. The middle document is a 'Closing Disclosure' with a title 'This form is a statement of loan terms and closing costs. Compare this document with your Loan Estimate.' The bottom document is a 'Uniform Residential Loan Application' form. It includes sections for 'I. TYPE OF MORTGAGE AND TERMS OF LOAN', 'II. PROPERTY INFORMATION AND PURPOSE OF LOAN', and 'III. BORROWER INFORMATION'. The form is filled out with sample data, such as 'Mortgage Applied for: Conventional', 'Amount: \$ 552,500', 'Interest Rate: 3.5 %', 'No. of Months: 360', 'Agency Case Number: ABC1234', 'Lender Case Number: XY25643', 'Subject Property Address: 123 Any Street, Anytown, USA, 12345', 'Borrower's Name: Carlos Salazar', and 'Co-Borrower's Name: N/A'.



Analyze Lending API

Classify & split mortgage documents

Extract data

Validate mortgage package



Automated end-to-end processing of mortgage packages



Pre-trained on diverse sets of mortgage document types



Faster time to value

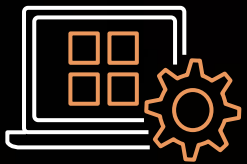
Mortgage loan package



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Analyze lending documents with a simple API

Upload a multi-page loan package



StartLendingAnalysis

Classify type of document



- Tax statement
- Wage statement
- Bank statement

Extract information specific to your business

Loan ID:	12345678
Loan type:	CONVENTIONAL
Loan amount:	\$552,500
Interest rate:	5.555%

Understand what information is present

Signature present on:

1005
check

Missing signature on:

1003

SPECIALIZED SUPPORT FOR MORTGAGE LOAN DOCUMENTS

OUTPUT



Amortization Type: FIXED RATE



Outputs normalized field names and supports implied elements

Automate mortgage document processing with AI/ML technologies





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



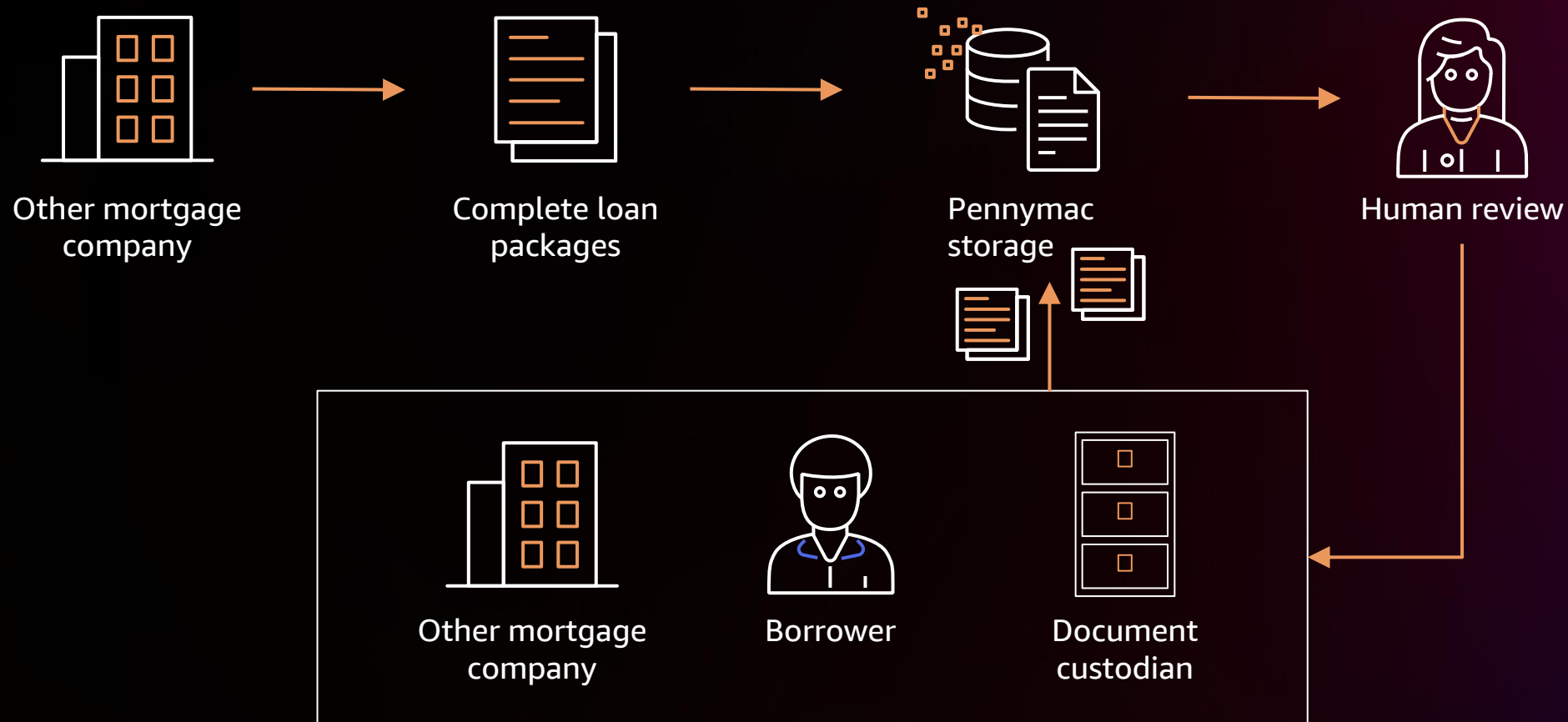
NMLS #35953

About PENNYMAC

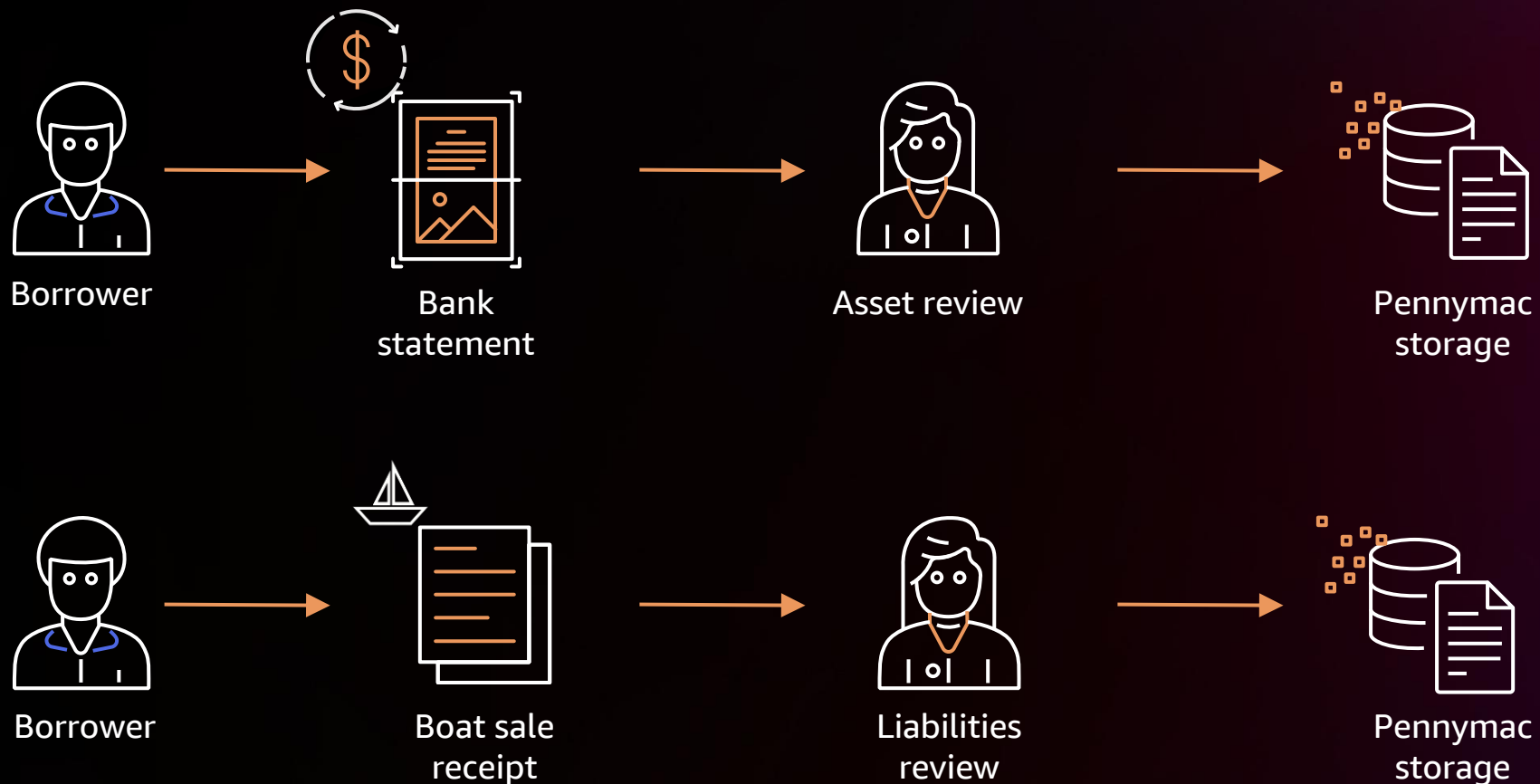
Company: As a top mortgage lender with over 4 million lifetime customers, Pennymac has a wide range of mortgage options available to suit the unique needs and financial goals of our current and prospective homeowners across the nation

Our mission is to build a foundation of homeownership by enabling people to achieve and sustain their aspirations of home

Processing – Correspondent lending



Processing – Direct lending



Millions of documents

Purchased loans

When loans are purchased, documents are delivered in a single PDF that's hundreds of pages long with no context or data



Created loans

When creating loans, context is fluid and documents are hard to organize

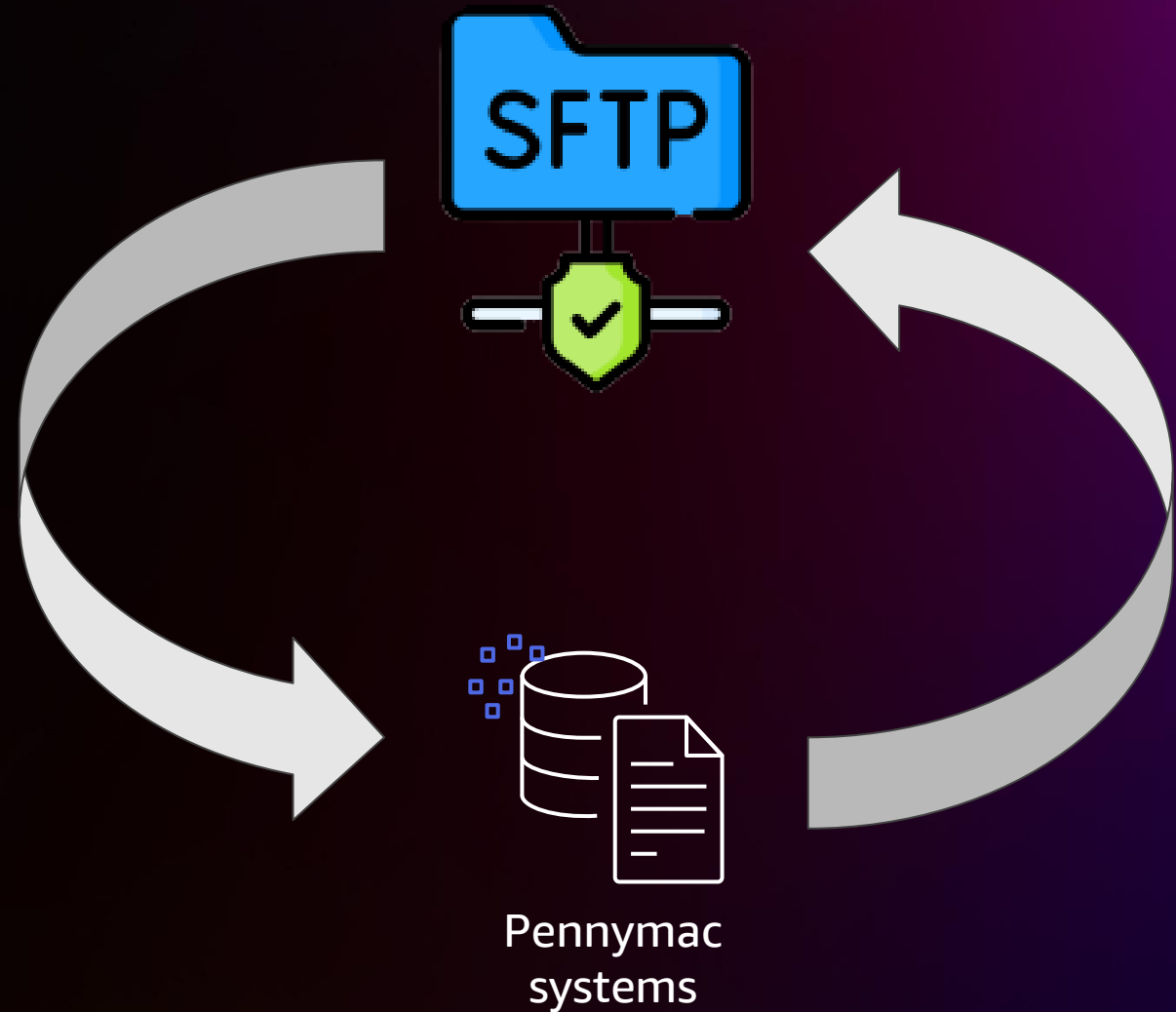


Massive cost

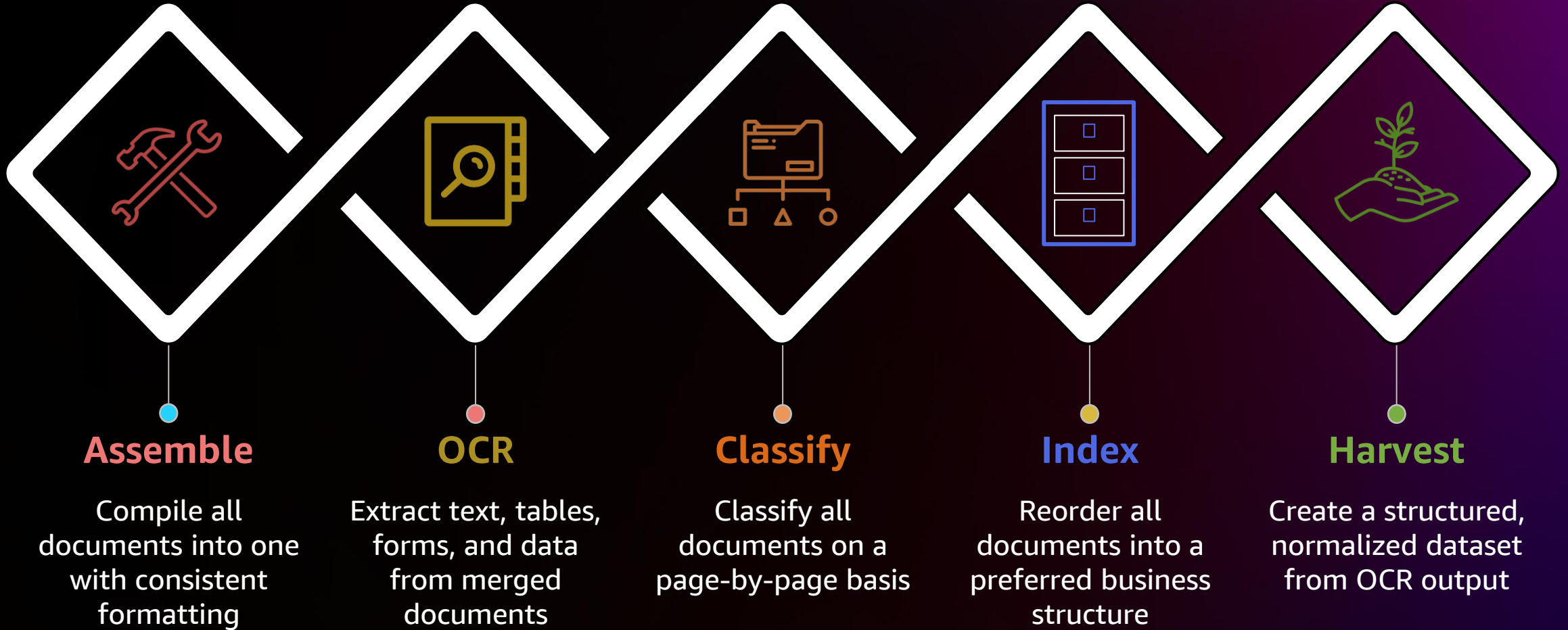
We typically spend over \$20 million annually in directly dealing with documents and lose out on much more in time spent

Vendor process

- ✗ No direct control of the process
- ✗ Inappropriate for multiple single file uploads
- ✗ No immediate feedback



Document processing at Pennymac



Amazon Textract – Benefits



Accuracy



Forms and tables

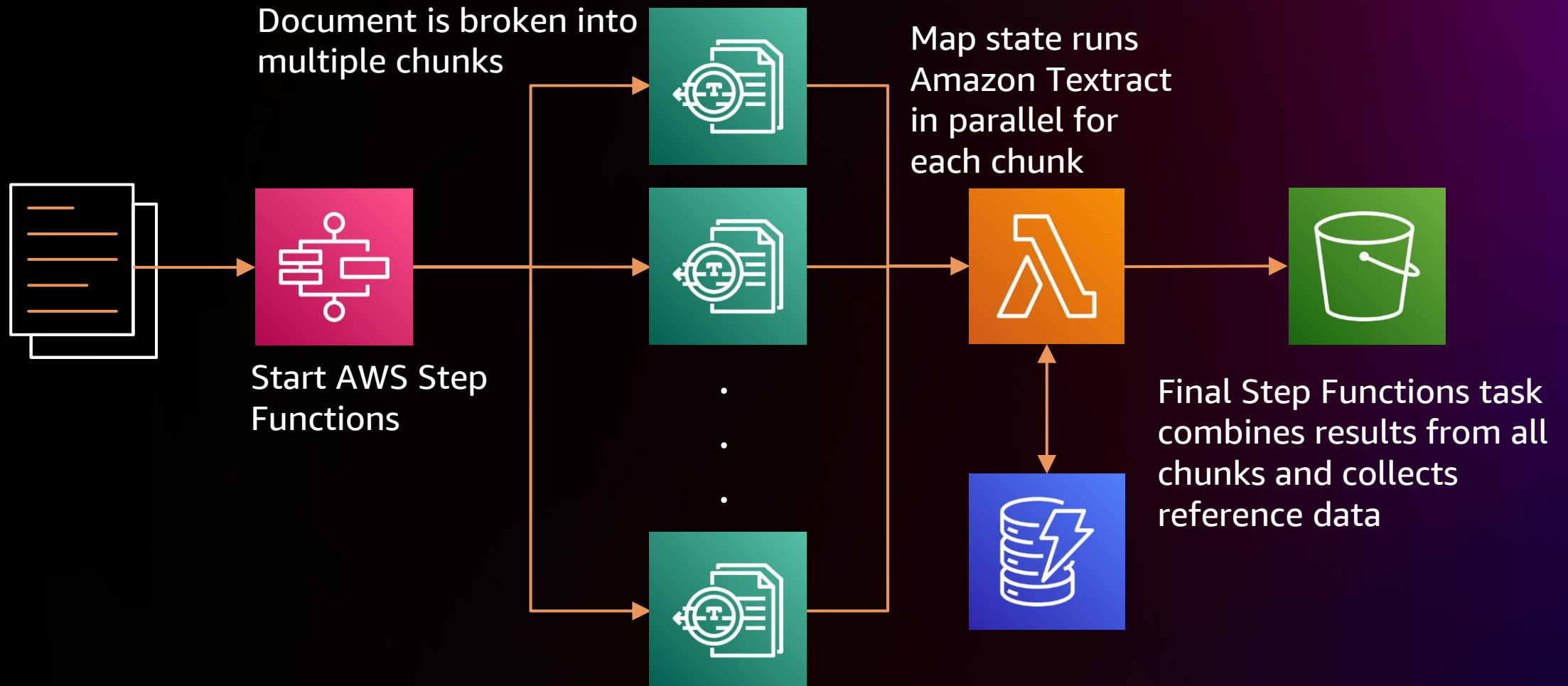


Queries



Helper libraries

Amazon Textract at scale



IDP with Amazon SageMaker

1

Collect

Our pipeline runs to aggregate OCR'd text data from our data lake into three separate data partitions

2

Tune

Pipeline takes 2 of the partitions for training and programmatically chooses different hyperparameter optimizations until the best permutation is found

3

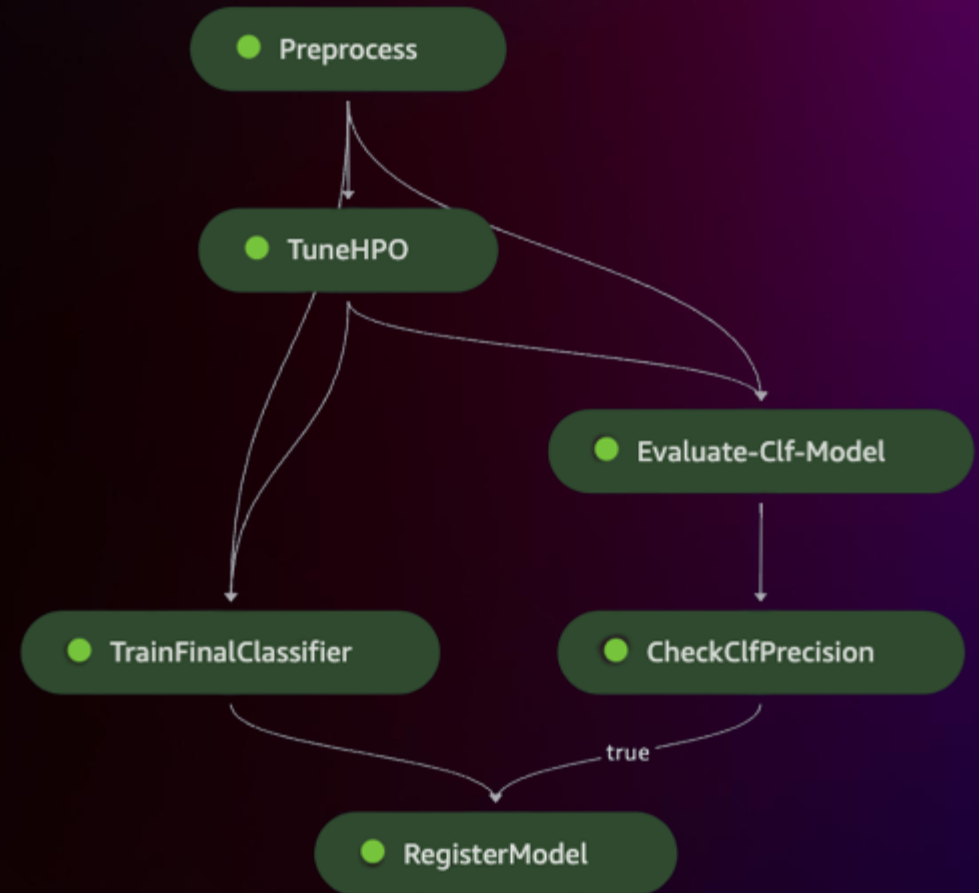
Evaluate

Remaining unseen data partition is used for testing the tuned model against; all evaluation metrics are logged and rendered in Amazon CloudWatch dashboards

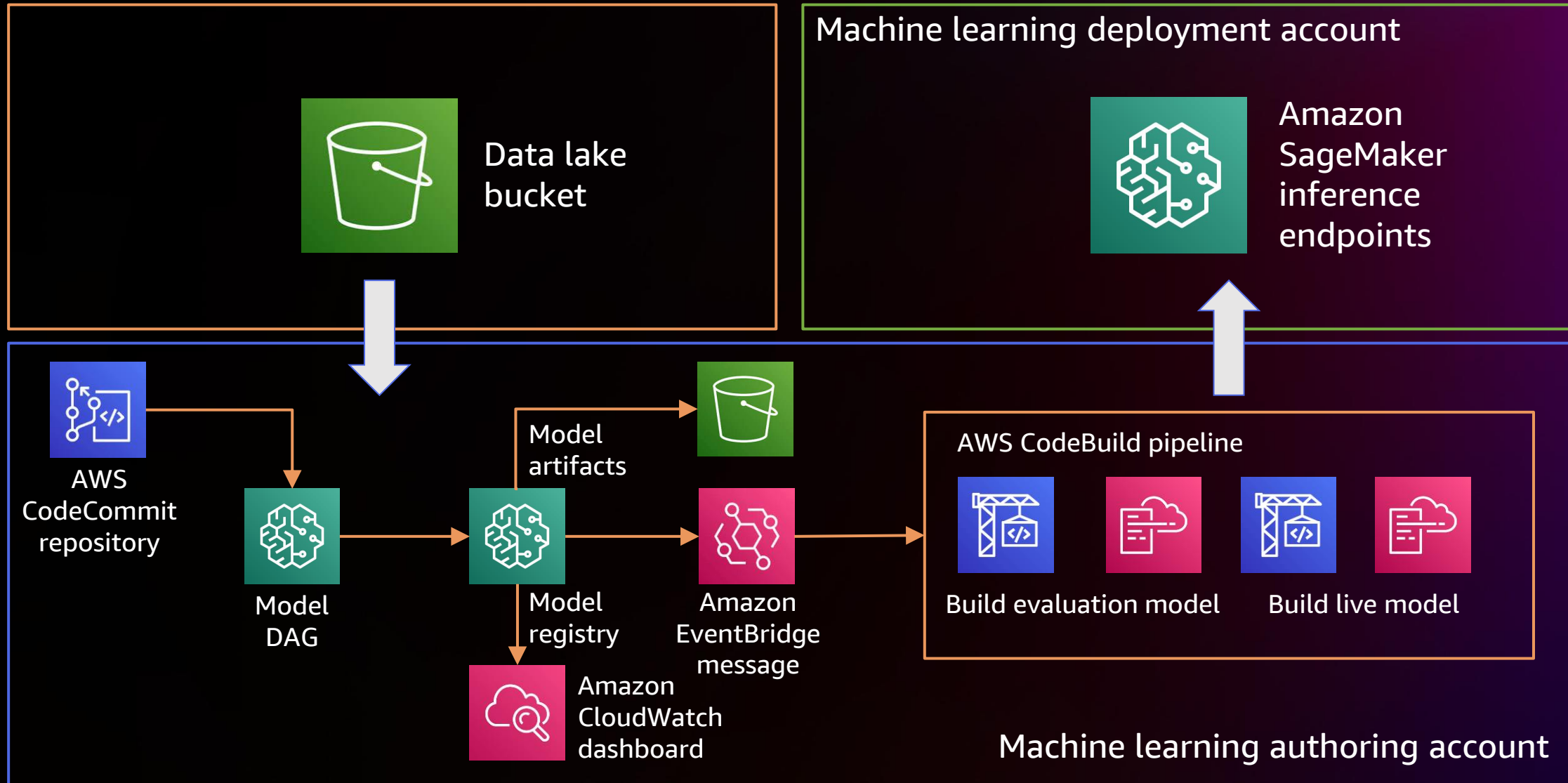
4

Train and register

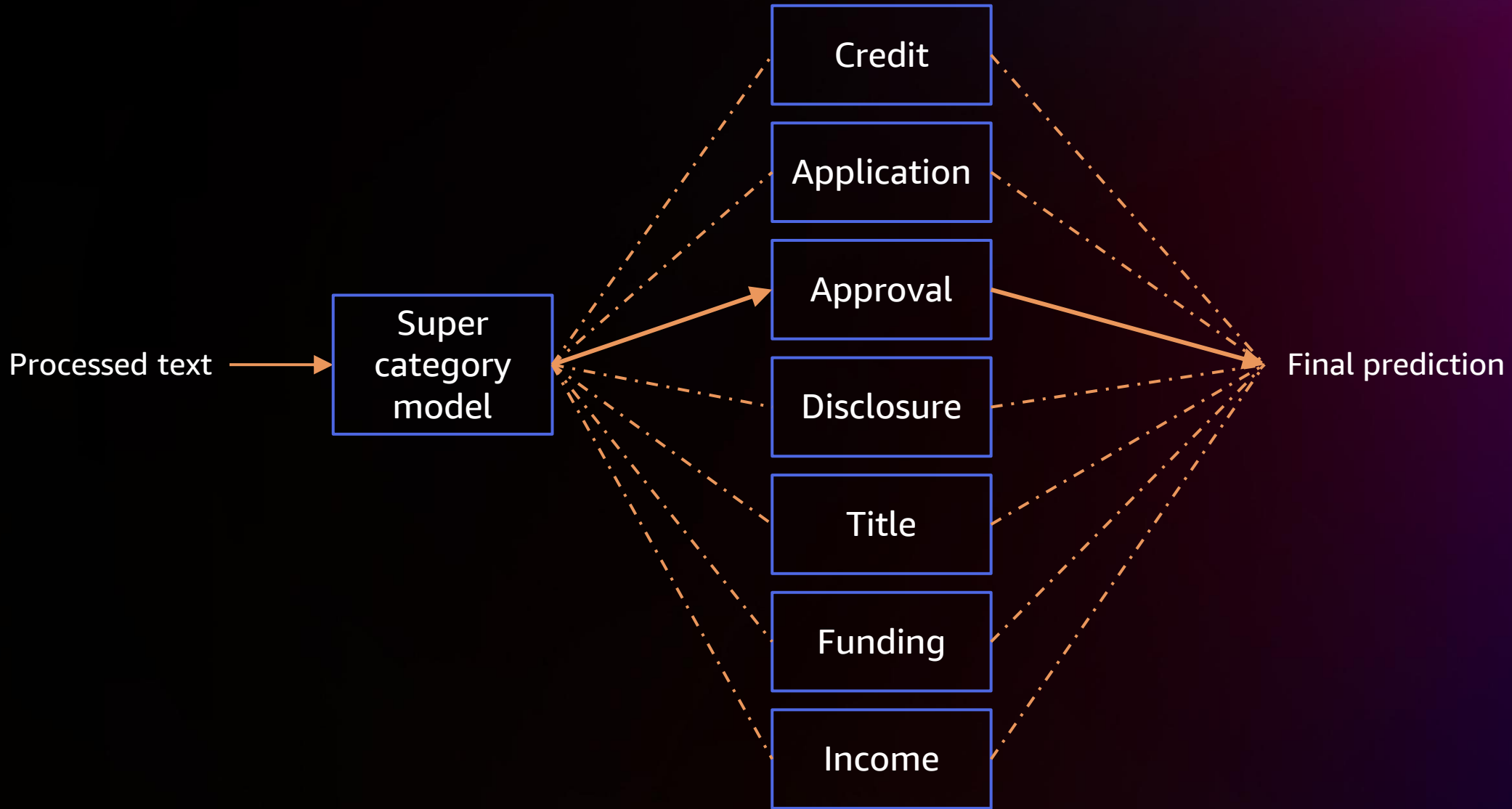
All data partitions are combined, and the model is trained on all data with the desired hyperparameters to expose the model to as much data as possible



Machine learning at Pennymac



Model inference for document classification



Data harvesting with GeoFinder

Borrower				Co-Borrower			
Borrower's Name (include Jr. or Sr. if applicable) Carlos Salazar				Co-Borrower's Name (include Jr. or Sr. if applicable) N/A			
Social Security Number 999-99-9999	Home Phone (incl. area code) +1 123-456-7890	DOB (mm/dd/yyyy) 1/18/1958	Yrs. School 12	Social Security Number N/A	Home Phone (incl. area code) N/A	DOB (mm/dd/yyyy) N/A	Yrs. School N/A
<input type="checkbox"/> Married (includes registered domestic partners) <input checked="" type="checkbox"/> Unmarried (includes single, divorced, widowed) <input type="checkbox"/> Separated		Dependents (not listed by Co-Borrower) No. 0 Ages N/A		<input type="checkbox"/> Married (includes registered domestic partners) <input type="checkbox"/> Unmarried (includes single, divorced, widowed) <input type="checkbox"/> Separated		Dependents (not listed by Borrower) No. N/A Ages N/A	
Present Address (street, city, state, ZIP/ country) <input type="checkbox"/> Own <input checked="" type="checkbox"/> Rent 5 No. Yrs 456 Any Street, Anytown, USA, 12345				Present Address (street, city, state, ZIP/ country) <input type="checkbox"/> Own <input type="checkbox"/> Rent ___ No. Yrs N/A			
Mailing Address, if different from Present Address				Mailing Address, if different from Present Address N/A			
<i>If residing at present address for less than two years, complete the following:</i>							
Former Address (street, city, state, ZIP) <input type="checkbox"/> Own <input type="checkbox"/> Rent ___ No. Yrs N/A				Former Address (street, city, state, ZIP) <input type="checkbox"/> Own <input type="checkbox"/> Rent ___ No. Yrs N/A			
Former Address (street, city, state, ZIP) <input type="checkbox"/> Own <input type="checkbox"/> Rent ___ No. Yrs N/A				Former Address (street, city, state, ZIP) <input type="checkbox"/> Own <input type="checkbox"/> Rent ___ No. Yrs N/A			

Uniform Residential Loan Application
Freddie Mac Form 65 7/05 (rev. 6/09)

Page 1

Borrower _____
Co-Borrower _____

Fannie Mae Form 1003 7/05 (rev. 6/09)

- Find form fields directly below given phrase
- Find form fields in between maximum and minimum X, Y coordinates
- Find given word or phrase on a desired page
- Get coordinates of lines

{ co_borrower_married: Not Selected,
co_borrower_unmarried: Not Selected,
co_borrower_separated: Not Selected }

<https://github.com/aws-samples/amazon-textract-textractor/tree/master/tpipelinegeofinder>

What's next?

Acceleration

Increase speed
to market for
data harvesting



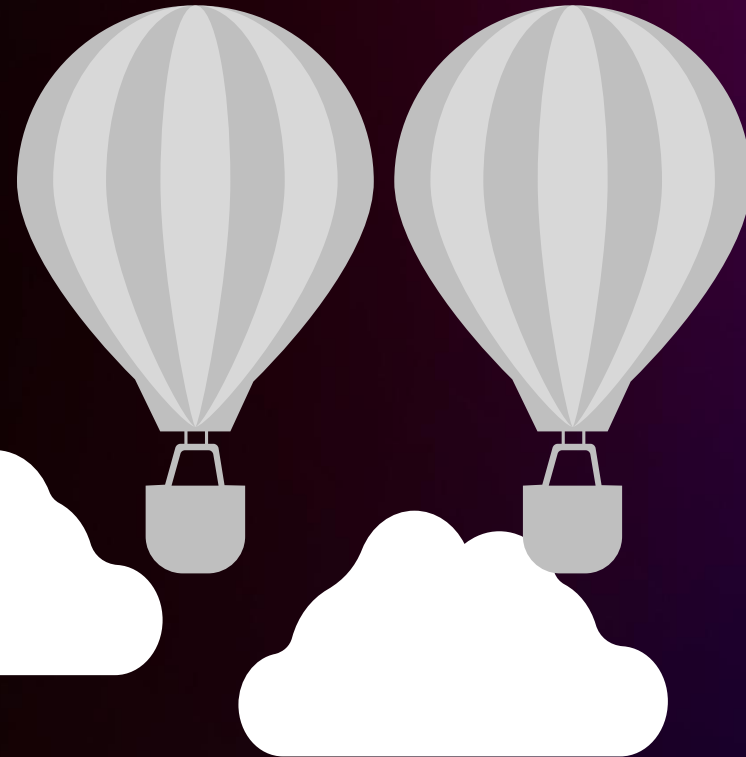
Less complex

A single endpoint to
receive normalized
data and document
classification



Normalized data

Remove the reliance on
expected phrases in
standardized
documents



AWS Partners can help accelerate IDP projects



	Suggest building in-house	Suggest AWS Partner
Document processing is a top corporate competency	X	
Difficulty hiring enough experienced AWS developers with knowledge of AI/ML		X
Using a partner for digital transformation projects		X
Already use serverless architecture to manage document workflows	X	



Important resources

✓ Service links

- Amazon Textract
bit.ly/3EN39Nh
- Amazon Comprehend
bit.ly/3gojWhn
- Amazon Augmented AI (Amazon A2I)
bit.ly/3GBn3wV

✓ Blog posts

- Build an end-to-end intelligent document processing solution using AWS
bit.ly/3goz7ab
- Automating customer onboarding with Lumiq's Drishti Document AI and Amazon Textract
bit.ly/3i3hqxu
- Broadridge Advances the Proxy Voting Data Process Using Artificial Intelligence on AWS
bit.ly/3VkdsPw

✓ Git source code link

- Document Understanding Solution
bit.ly/3UYVMcd
- Amazon Textract samples
bit.ly/3go0ykD
- Amazon Comprehend samples
bit.ly/3XpPtQw
- Amazon A2I samples
bit.ly/3Osrj3J

✓ Account team can provide

- Amazon Textract and Amazon Comprehend security papers (can be shared under NDA)
- Immersion day training
- AWS Partner referrals

Thank you!



Please complete the session survey in the **mobile app**

