aws INNOVATE

AI/ML EDITION

aws INNOVATE

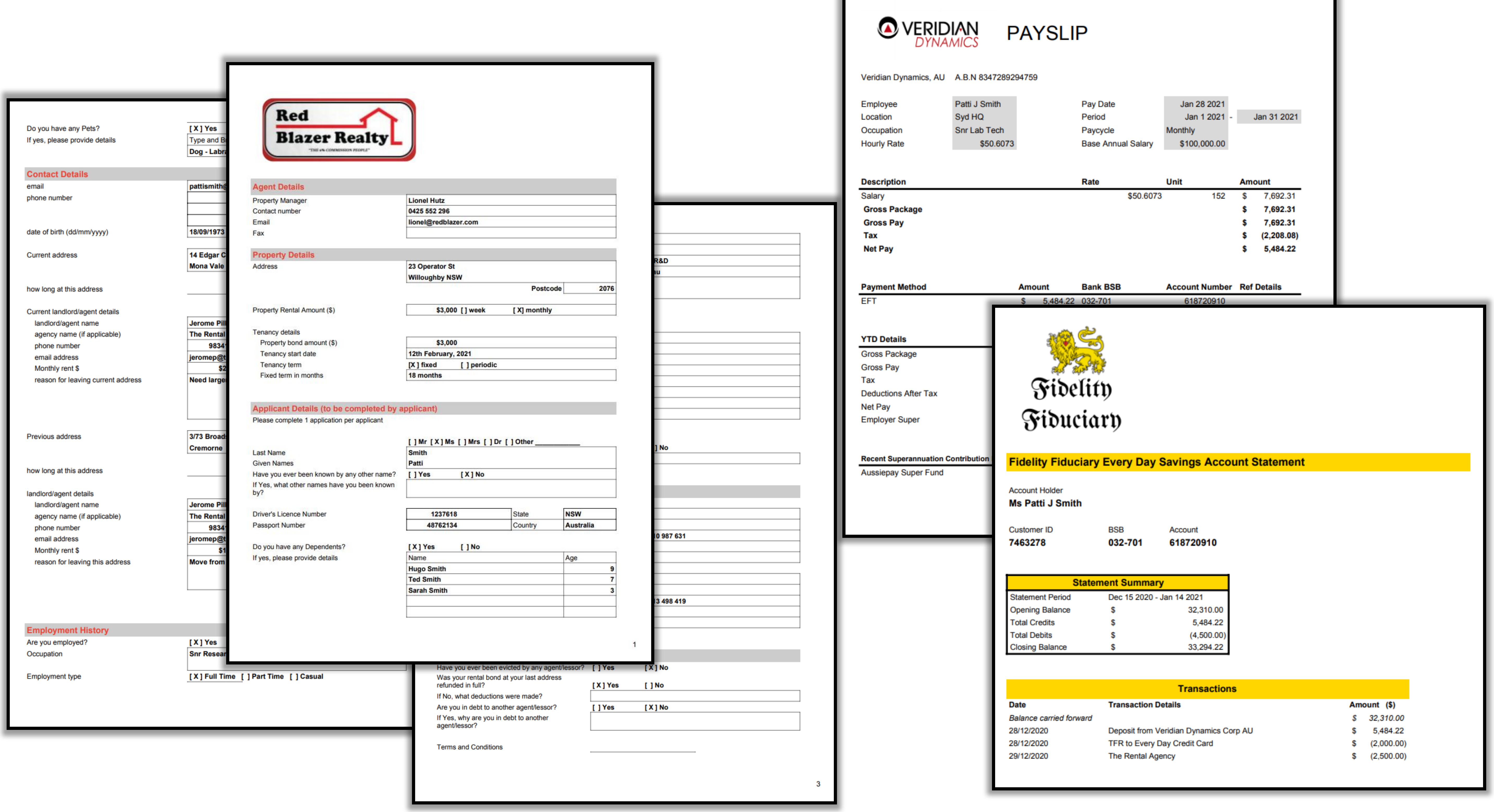
AI/ML EDITION

Intelligent document processing: Building higher accuracy document automation at scale

Jonathan Hedley
Principal Solutions Architect
Amazon Web Services

Agenda

- Building an automated, event driven document processing system
- Use case
- Architecture
- Code samples
- Demo



Do you have any Pets?	[X] Yes
If yes, please provide details	Type and B
	Dog - Labra

Contact Details

email	pattismith@
phone number	
date of birth (dd/mm/yyyy)	18/09/1973
Current address	14 Edgar C Mona Vale
how long at this address	
Current landlord/agent details	
landlord/agent name	Jerome Pil
agency name (if applicable)	The Rental
phone number	9834
email address	jeromep@t
Monthly rent $	$2
reason for leaving current address	Need large
Previous address	3/73 Broad Cremorne
how long at this address	
landlord/agent details	
landlord/agent name	Jerome Pil
agency name (if applicable)	The Rental
phone number	9834
email address	jeromep@t
Monthly rent $	$1
reason for leaving this address	Move from

Employment History

Are you employed?	[X] Yes
Occupation	Snr Resear
Employment type	[X] Full Time [ ] Part Time [ ] Casual

Red Blazer Realty

"THE 6% COMMISSION PEOPLE"

Agent Details

Property Manager	Lionel Hutz
Contact number	0425 552 296
Email	lionel@redblazer.com
Fax	

Property Details

Address	23 Operator St Willoughby NSW		
		Postcode	2076
Property Rental Amount ($)	$3,000 [ ] week [X] monthly		
Tenancy details			
Property bond amount ($)	$3,000		
Tenancy start date	12th February, 2021		
Tenancy term	[X] fixed [ ] periodic		
Fixed term in months	18 months		

Applicant Details (to be completed by applicant)

Please complete 1 application per applicant

	[ ] Mr [X] Ms [ ] Mrs [ ] Dr [ ] Other ___________		
Last Name	Smith		
Given Names	Patti		
Have you ever been known by any other name?	[ ] Yes [X] No		
If Yes, what other names have you been known by?			
Driver's Licence Number	1237618	State	NSW
Passport Number	48762134	Country	Australia
Do you have any Dependents?	[X] Yes [ ] No		
If yes, please provide details	Name	Age	
	Hugo Smith	9	
	Ted Smith	7	
	Sarah Smith	3	

1

R&D

au

] No

0 987 631

3 498 419

Have you ever been evicted by any agent/lessor?	[ ] Yes	[X] No
Was your rental bond at your last address refunded in full?	[X] Yes	[ ] No
If No, what deductions were made?		
Are you in debt to another agent/lessor?	[ ] Yes	[X] No
If Yes, why are you in debt to another agent/lessor?		
Terms and Conditions		

3

VERIDIAN DYNAMICS

PAYSLIP

Veridian Dynamics, AU A.B.N 8347289294759

Employee	Patti J Smith	Pay Date	Jan 28 2021	
Location	Syd HQ	Period	Jan 1 2021 -	Jan 31 2021
Occupation	Snr Lab Tech	Paycycle	Monthly	
Hourly Rate	$50.6073	Base Annual Salary	$100,000.00	

Description	Rate	Unit	Amount	
Salary	$50.6073	152	$	7,692.31
Gross Package			**$**	**7,692.31**
Gross Pay			**$**	**7,692.31**
Tax			**$**	**(2,208.08)**
Net Pay			**$**	**5,484.22**

Payment Method	Amount	Bank BSB	Account Number	Ref Details
EFT	$ 5,484.22	032-701	618720910	

YTD Details

Gross Package

Gross Pay

Tax

Deductions After Tax

Net Pay

Employer Super

Recent Superannuation Contribution

Aussiepay Super Fund

Fidelity Fiduciary

Fidelity Fiduciary Every Day Savings Account Statement

Account Holder

Ms Patti J Smith

Customer ID	BSB	Account
7463278	**032-701**	**618720910**

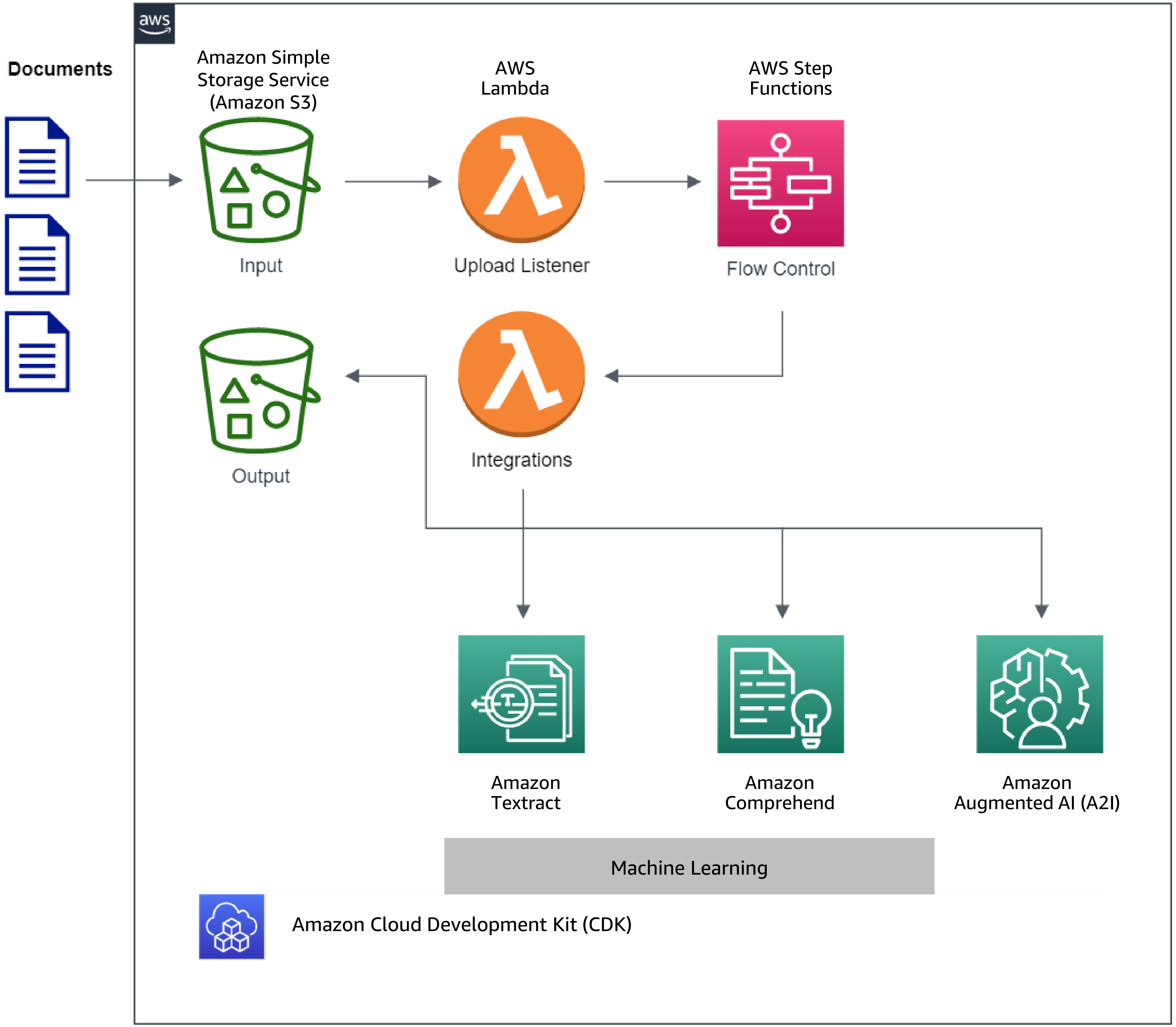
Statement Summary

Statement Period	Dec 15 2020 - Jan 14 2021	
Opening Balance	$	32,310.00
Total Credits	$	5,484.22
Total Debits	$	(4,500.00)
Closing Balance	$	33,294.22

Transactions

Date	Transaction Details	Amount ($)	
Balance carried forward		*$*	*32,310.00*
28/12/2020	Deposit from Veridian Dynamics Corp AU	$	5,484.22
28/12/2020	TFR to Every Day Credit Card	$	(2,000.00)
29/12/2020	The Rental Agency	$	(2,500.00)

Architecture

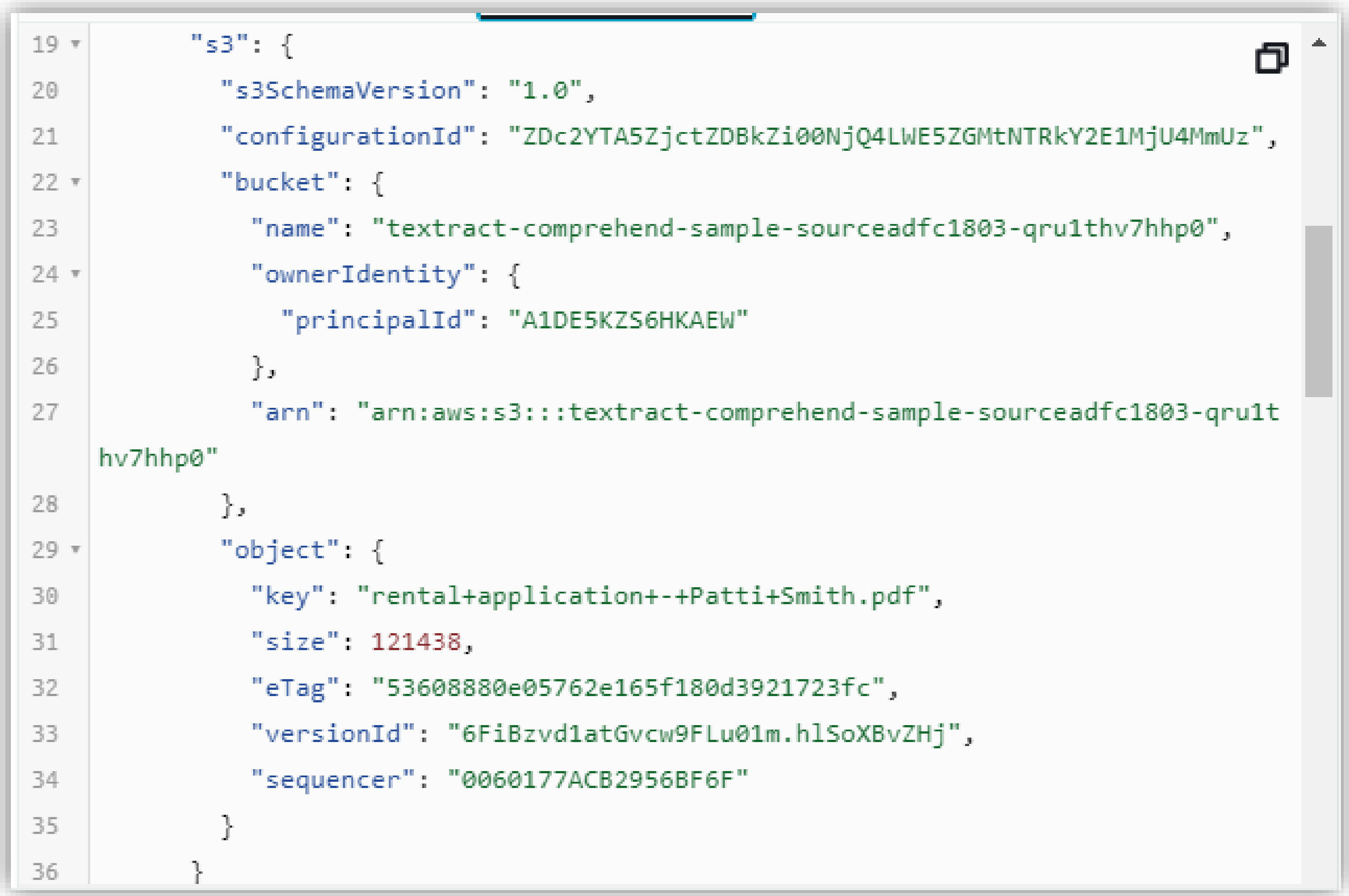


Amazon S3 upload event to AWS Lambda

Amazon Simple Storage Service (Amazon S3)

Source bucket

Output bucket

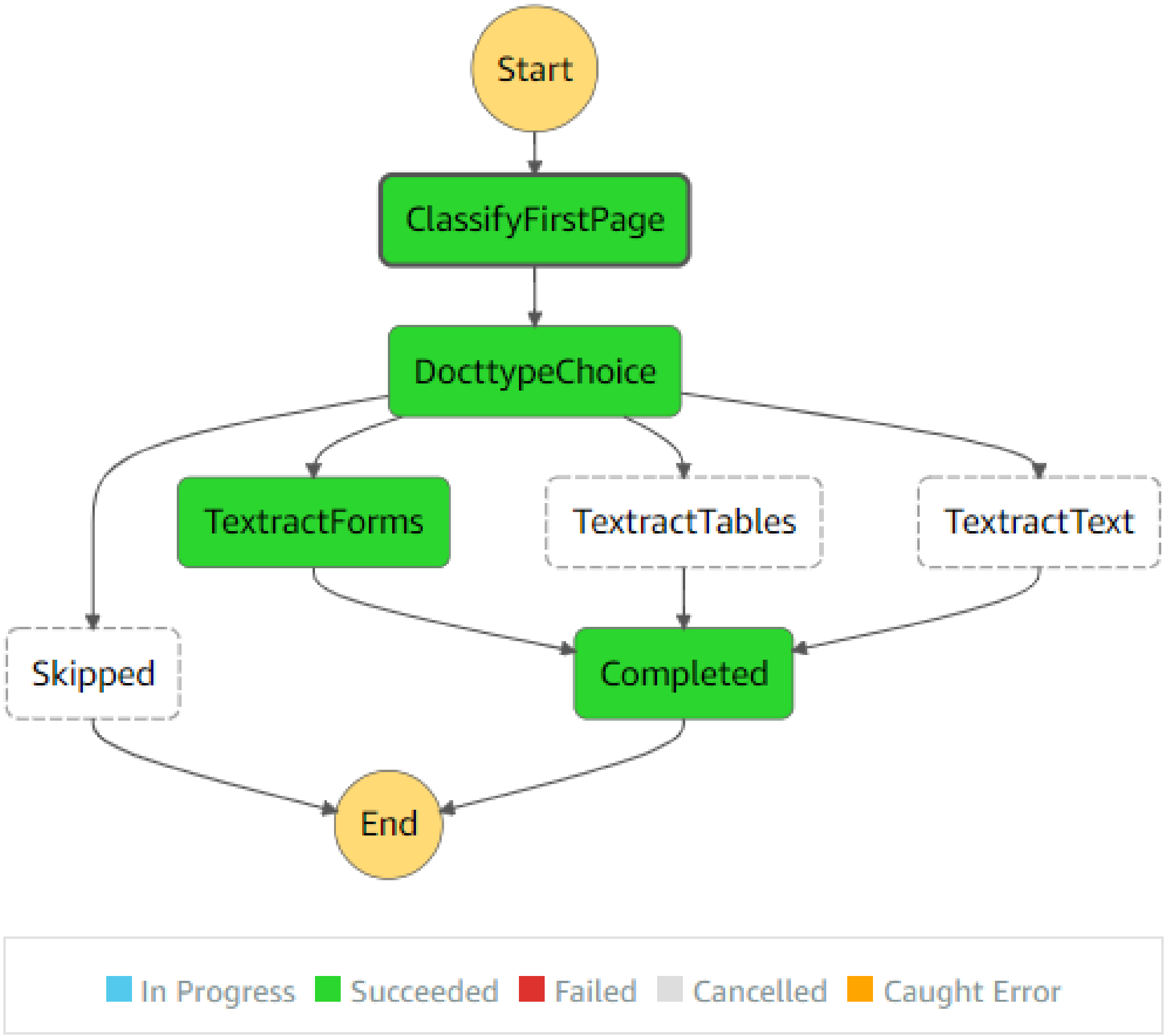


```
      "s3": {
        "s3SchemaVersion": "1.0",
        "configurationId": "ZDc2YTA5ZjctZDBkZi00NjQ4LWE5ZGMtNTRkY2E1MjU4MmUz",
        "bucket": {
          "name": "textract-comprehend-sample-sourceadfc1803-qru1thv7hhp0",
          "ownerIdentity": {
            "principalId": "A1DE5KZS6HKAEW"
          },
          "arn": "arn:aws:s3:::textract-comprehend-sample-sourceadfc1803-qru1t
hv7hhp0"
        },
        "object": {
          "key": "rental+application+-+Patti+Smith.pdf",
          "size": 121438,
          "eTag": "53608880e05762e165f180d3921723fc",
          "versionId": "6FiBzvd1atGvcw9FLu01m.hlSoXBvZHj",
          "sequencer": "0060177ACB2956BF6F"
        }
      }
```

AWS Lambda

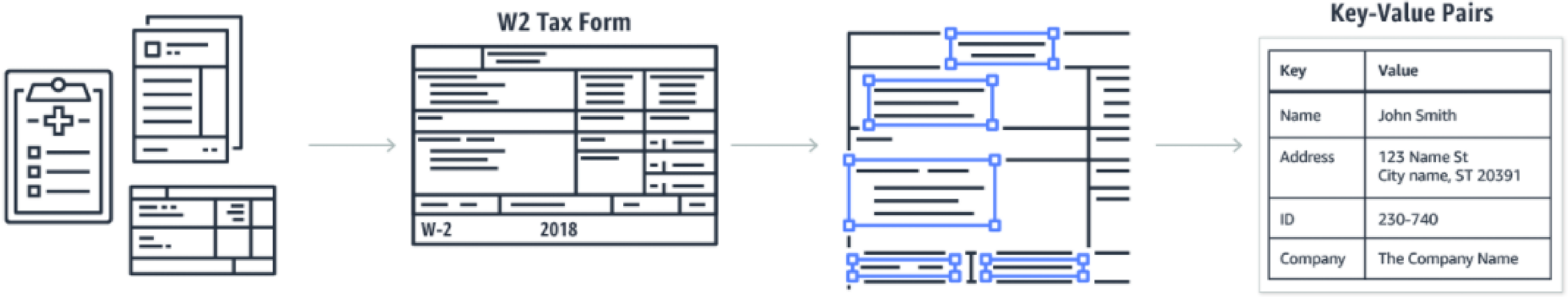
AWS Step Functions flow

AWS Step Functions



Amazon Textract

Amazon Textract



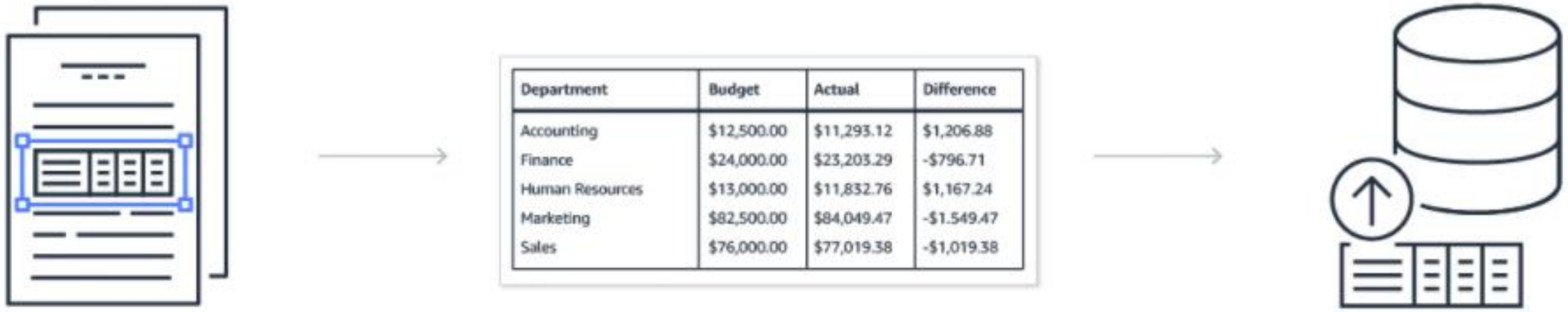
Key	Value
Name	John Smith
Address	123 Name St City name, ST 20391
ID	230-740
Company	The Company Name

Tax, medical, banking, and other form documents

Textract recognizes many forms, such as W2, 1099-MISC, 1040, patient registration, and more

Automatically process documents without data entry or writing extraction rules

Automatically extract key-value pairs and retain document context without manual intervention



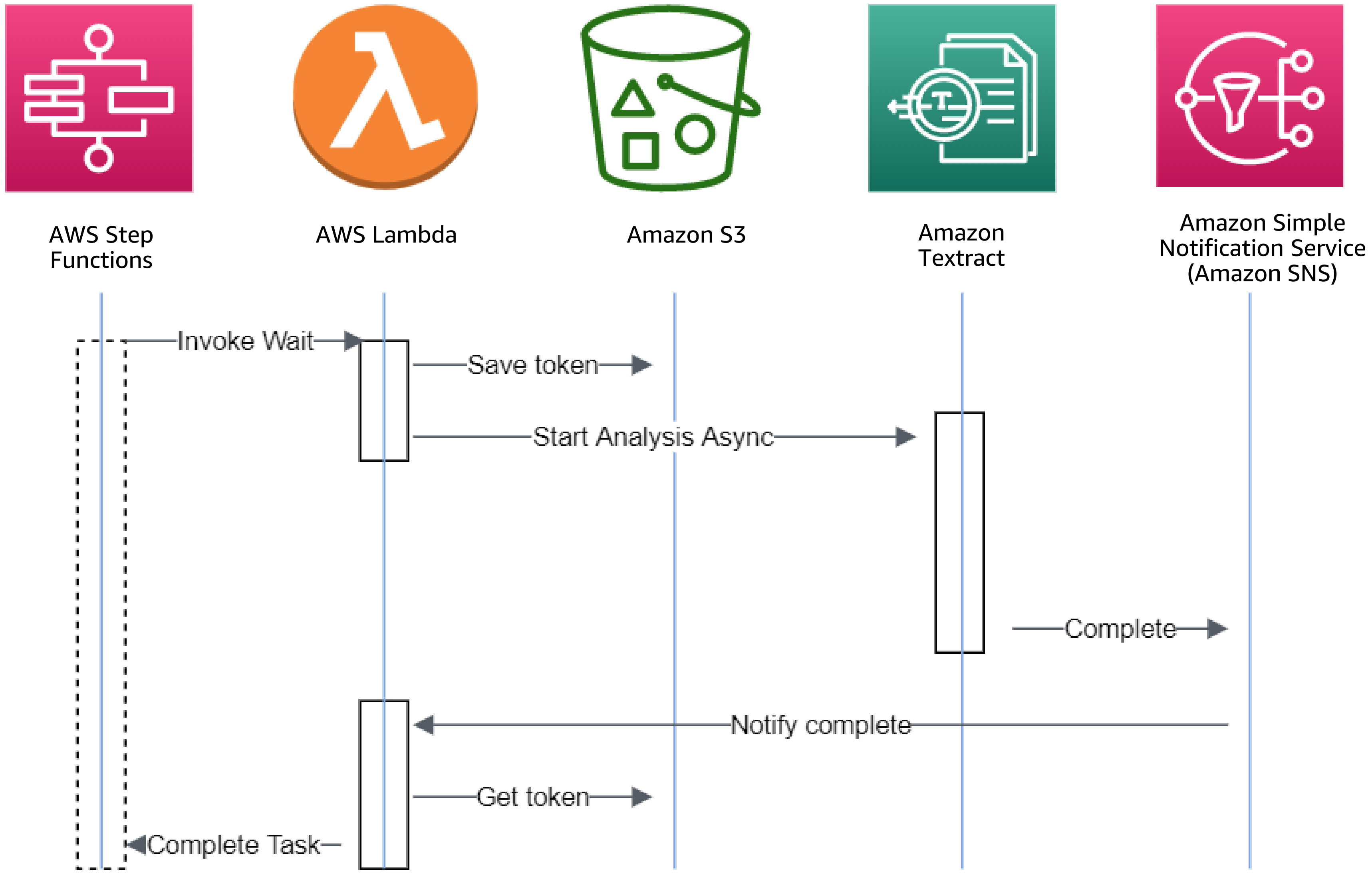
Department	Budget	Actual	Difference
Accounting	$12,500.00	$11,293.12	$1,206.88
Finance	$24,000.00	$23,203.29	-$796.71
Human Resources	$13,000.00	$11,832.76	$1,167.24
Marketing	$82,500.00	$84,049.47	-$1.549.47
Sales	$76,000.00	$77,019.38	-$1,019.38

When extracting text from documents and forms, Textract automatically detects and extracts structured data

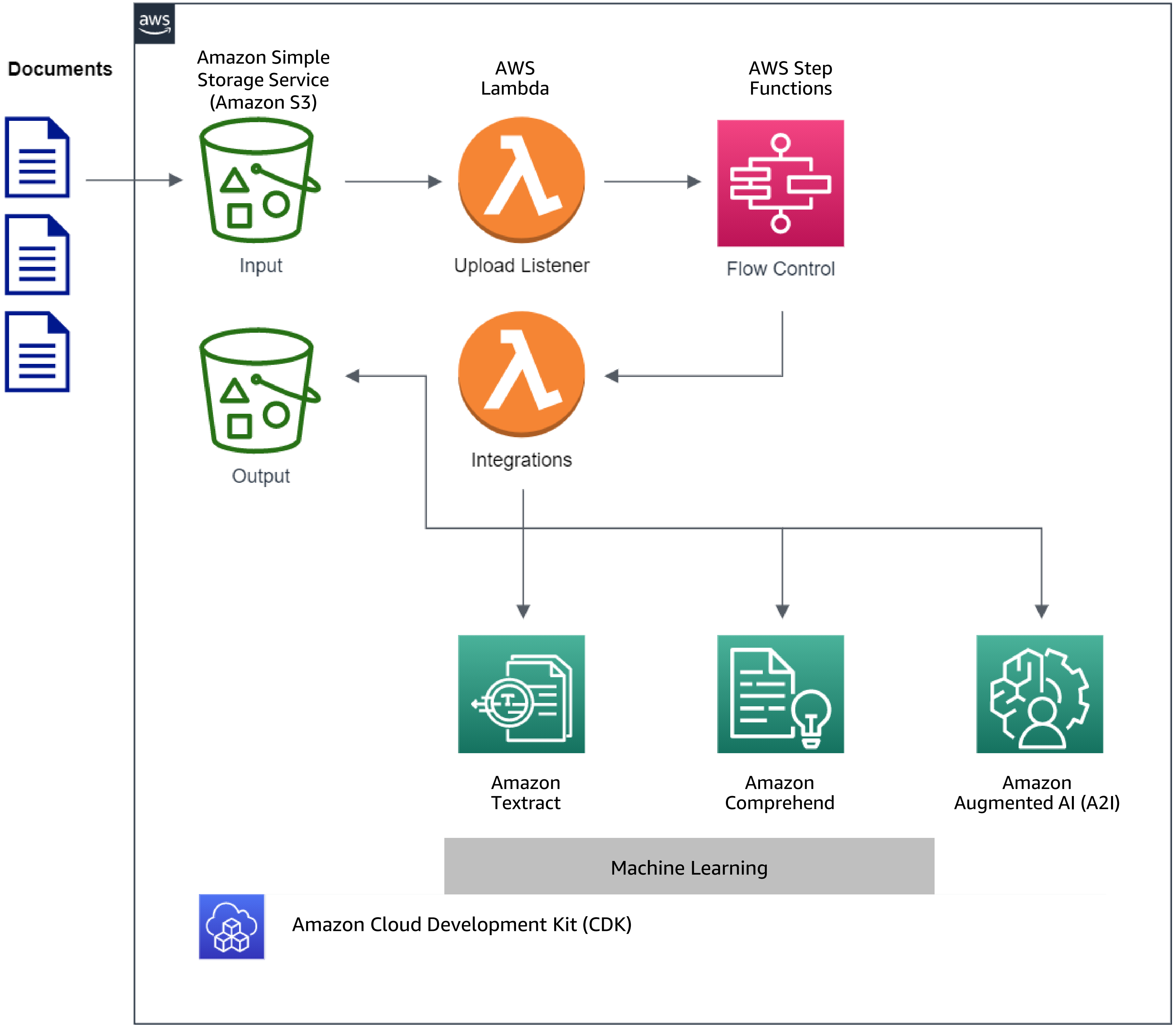
Textract preserves the tabular structure of extracted data, so that text remains grouped within each cell

With the tabular format of the data intact, easily upload extracted data into a database

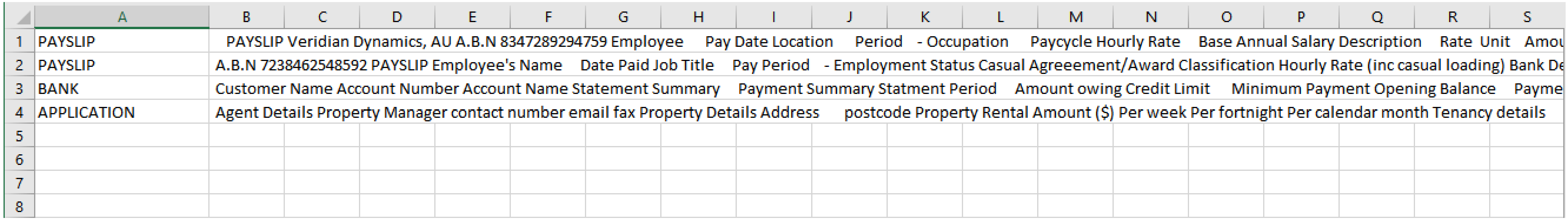
AWS Step Functions callback pattern



Architecture

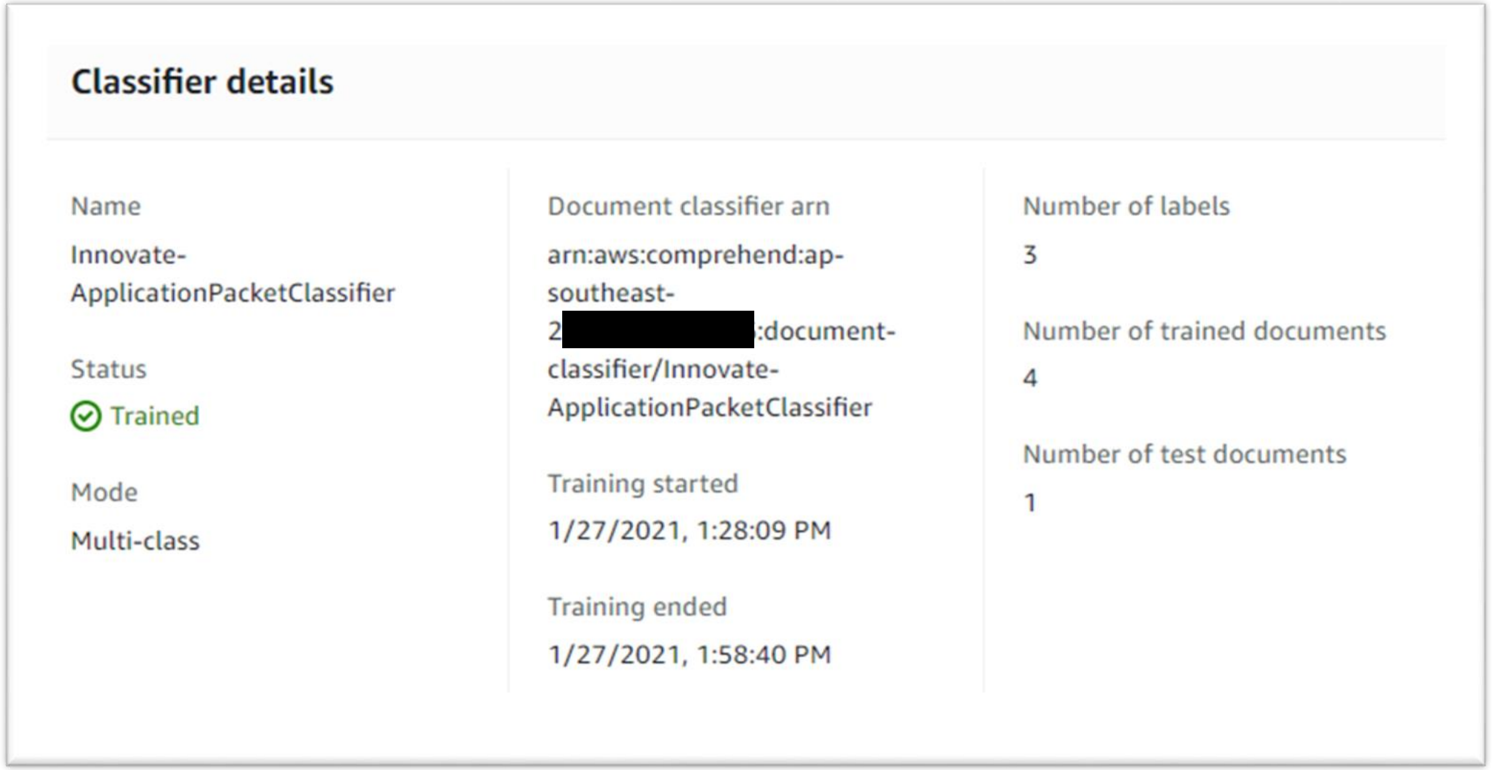


Amazon Comprehend classifier



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	PAYSLIP	PAYSLIP Veridian Dynamics, AU A.B.N 8347289294759 Employee Pay Date Location Period - Occupation Paycycle Hourly Rate Base Annual Salary Description Rate Unit Amou																	
2	PAYSLIP	A.B.N 7238462548592 PAYSLIP Employee's Name Date Paid Job Title Pay Period - Employment Status Casual Agreeement/Award Classification Hourly Rate (inc casual loading) Bank De																	
3	BANK	Customer Name Account Number Account Name Statement Summary Payment Summary Statment Period Amount owing Credit Limit Minimum Payment Opening Balance Payme																	
4	APPLICATION	Agent Details Property Manager contact number email fax Property Details Address postcode Property Rental Amount ($) Per week Per fortnight Per calendar month Tenancy details																	
5																			
6																			
7																			
8																			

Amazon Comprehend



Classifier details

Name
Innovate-ApplicationPacketClassifier

Status
✓ Trained

Mode
Multi-class

Document classifier arn
arn:aws:comprehend:ap-southeast-2[redacted]:document-classifier/Innovate-ApplicationPacketClassifier

Training started
1/27/2021, 1:28:09 PM

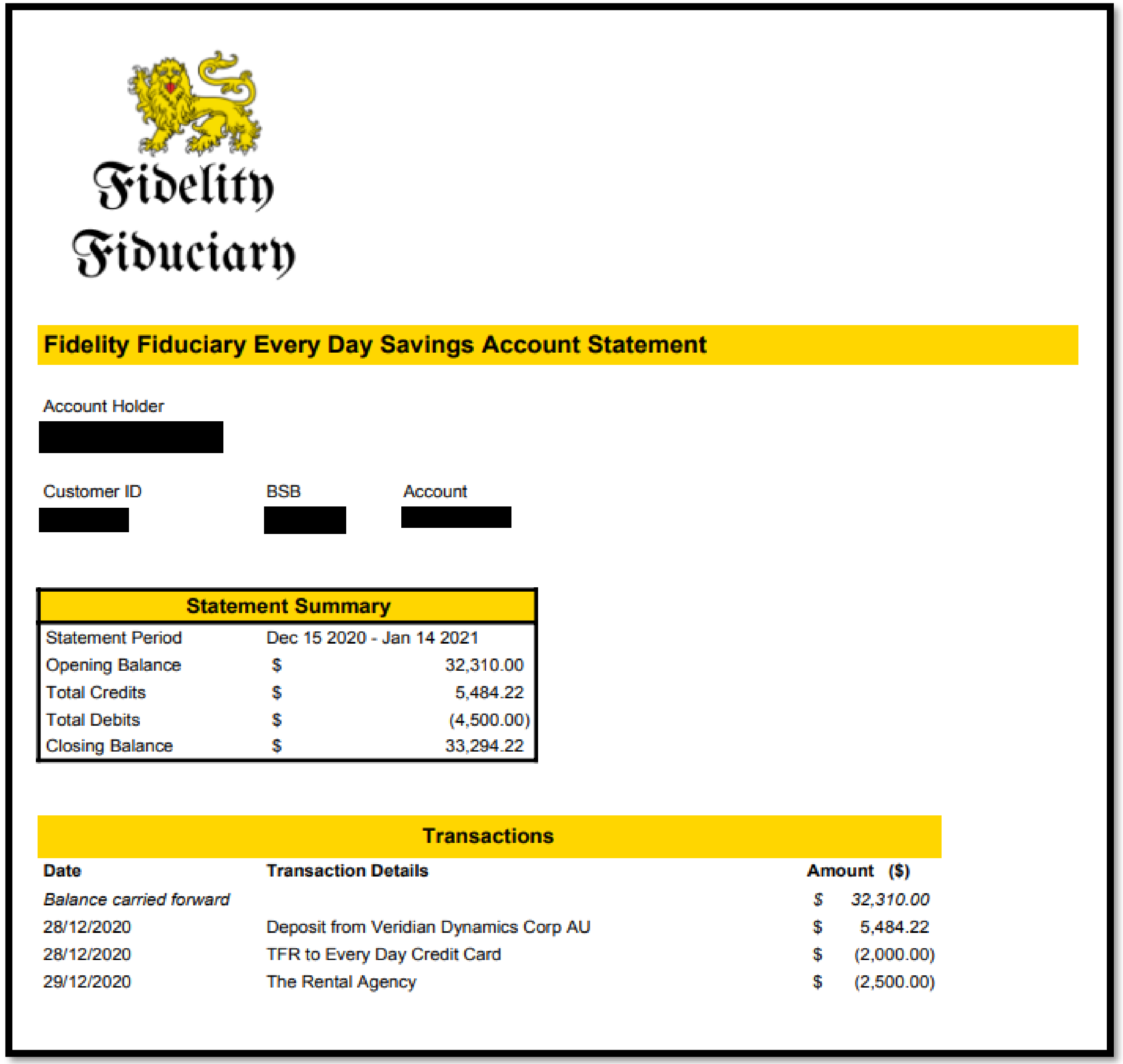
Training ended
1/27/2021, 1:58:40 PM

Number of labels
3

Number of trained documents
4

Number of test documents
1

Redaction



Fidelity Fiduciary

Fidelity Fiduciary Every Day Savings Account Statement

Account Holder

[redacted]

Customer ID	BSB	Account
[redacted]	[redacted]	[redacted]

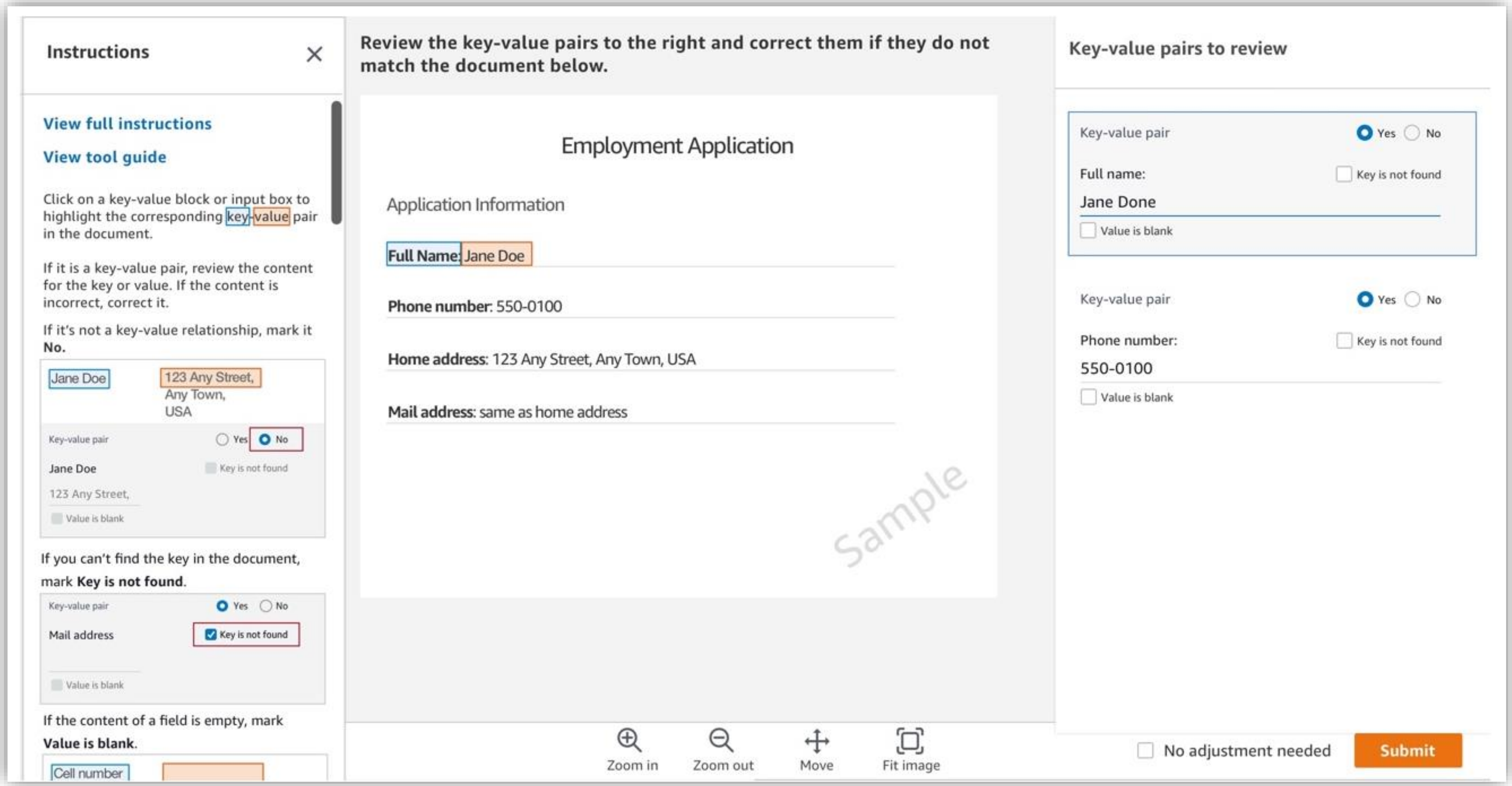
Statement Summary		
Statement Period	Dec 15 2020 - Jan 14 2021	
Opening Balance	$	32,310.00
Total Credits	$	5,484.22
Total Debits	$	(4,500.00)
Closing Balance	$	33,294.22

Transactions

Date	**Transaction Details**	**Amount ($)**
Balance carried forward		*$ 32,310.00*
28/12/2020	Deposit from Veridian Dynamics Corp AU	$ 5,484.22
28/12/2020	TFR to Every Day Credit Card	$ (2,000.00)
29/12/2020	The Rental Agency	$ (2,500.00)

Amazon Augmented AI (A2I)

Amazon Augmented AI (A2I)



Instructions ✕

View full instructions

View tool guide

Click on a key-value block or input box to highlight the corresponding key-value pair in the document.

If it is a key-value pair, review the content for the key or value. If the content is incorrect, correct it.

If it's not a key-value relationship, mark it **No.**

Jane Doe 123 Any Street, Any Town, USA

Key-value pair ○ Yes ● No

Jane Doe ☐ Key is not found

123 Any Street,

☐ Value is blank

If you can't find the key in the document, mark **Key is not found.**

Key-value pair ● Yes ○ No

Mail address ☑ Key is not found

☐ Value is blank

If the content of a field is empty, mark **Value is blank.**

Cell number

Review the key-value pairs to the right and correct them if they do not match the document below.

Employment Application

Application Information

Full Name: Jane Doe

Phone number: 550-0100

Home address: 123 Any Street, Any Town, USA

Mail address: same as home address

Key-value pairs to review

Key-value pair ● Yes ○ No

Full name: ☐ Key is not found

Jane Done

☐ Value is blank

Key-value pair ● Yes ○ No

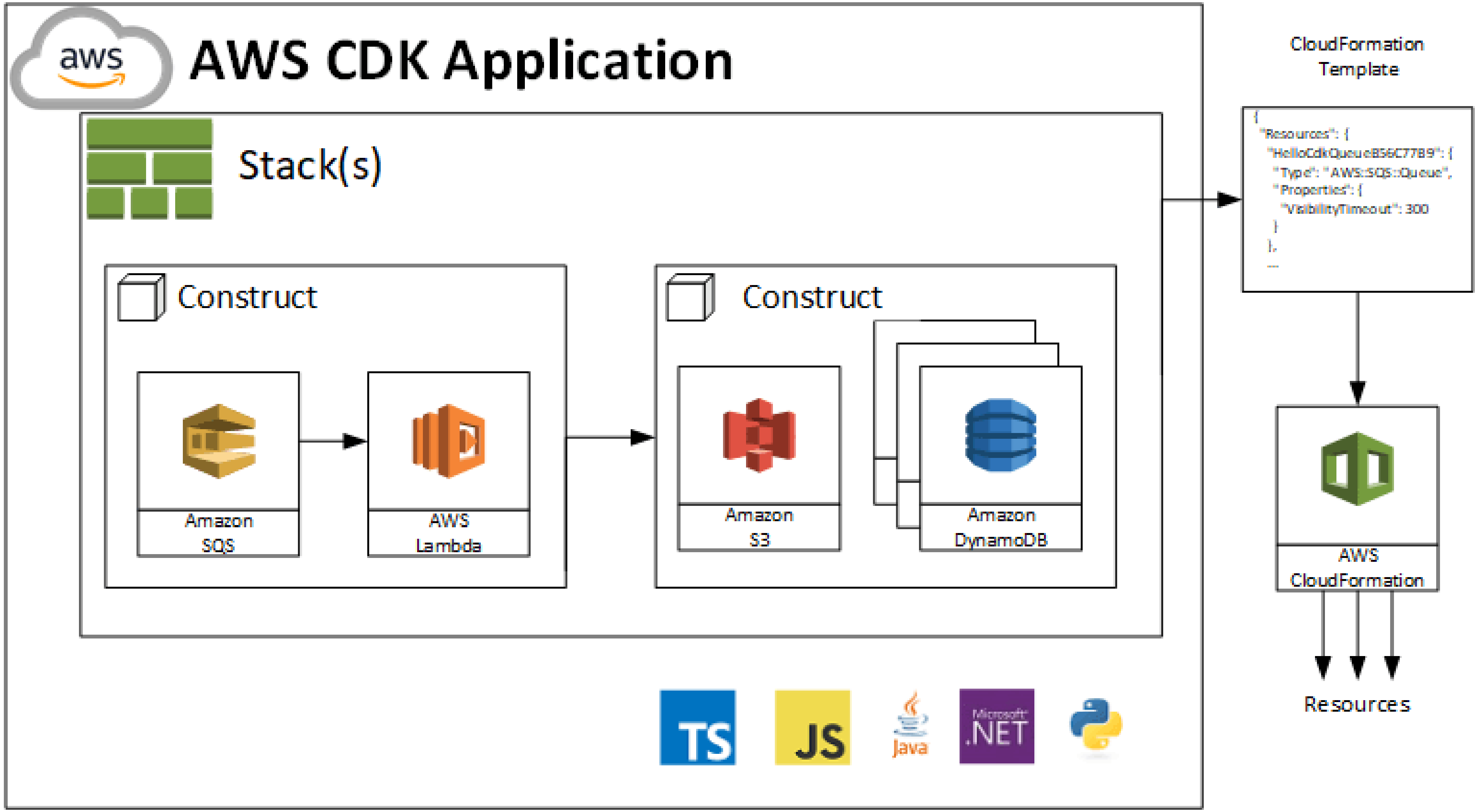
Phone number: ☐ Key is not found

550-0100

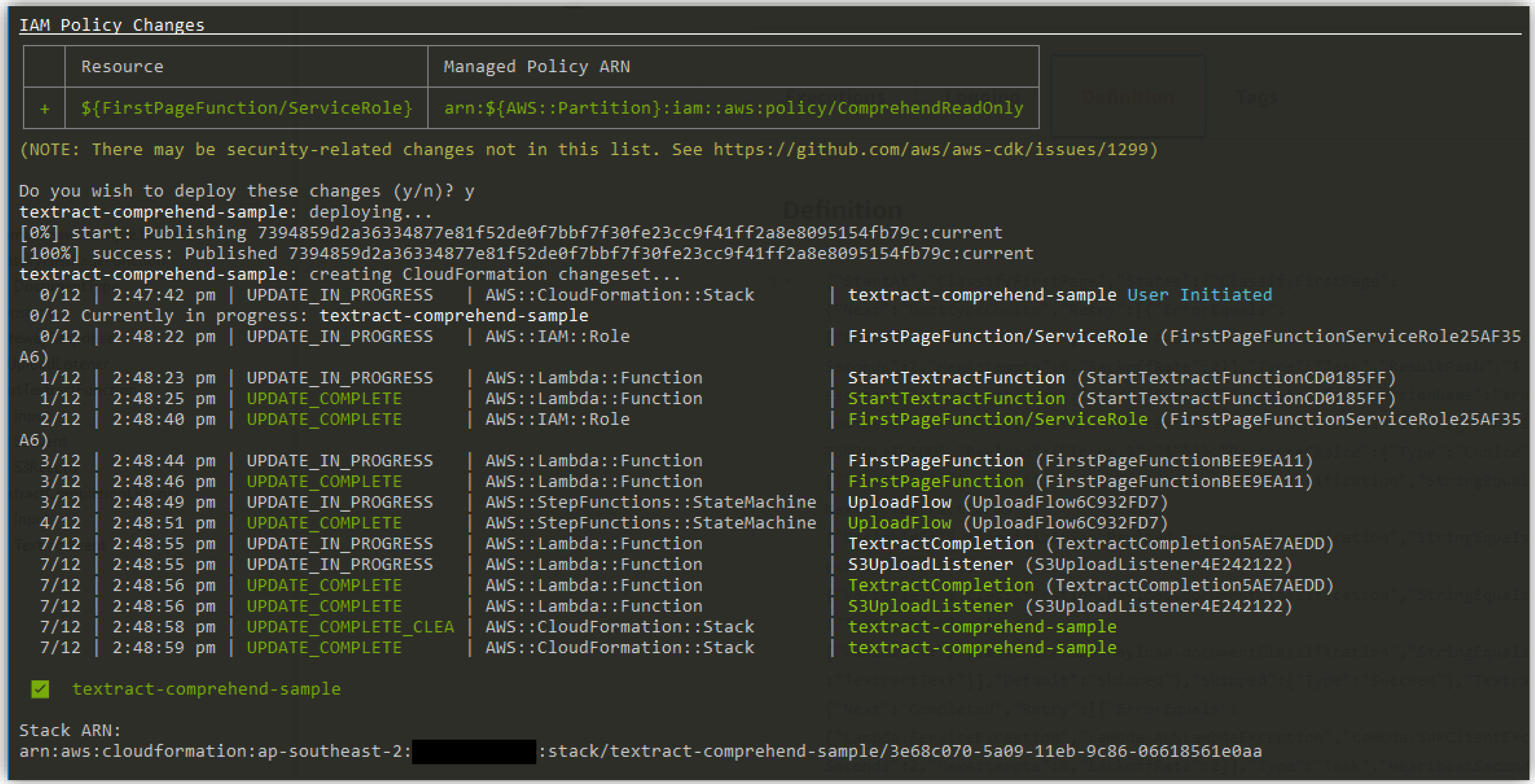
☐ Value is blank

☐ No adjustment needed

AWS Cloud Development Kit (CDK)

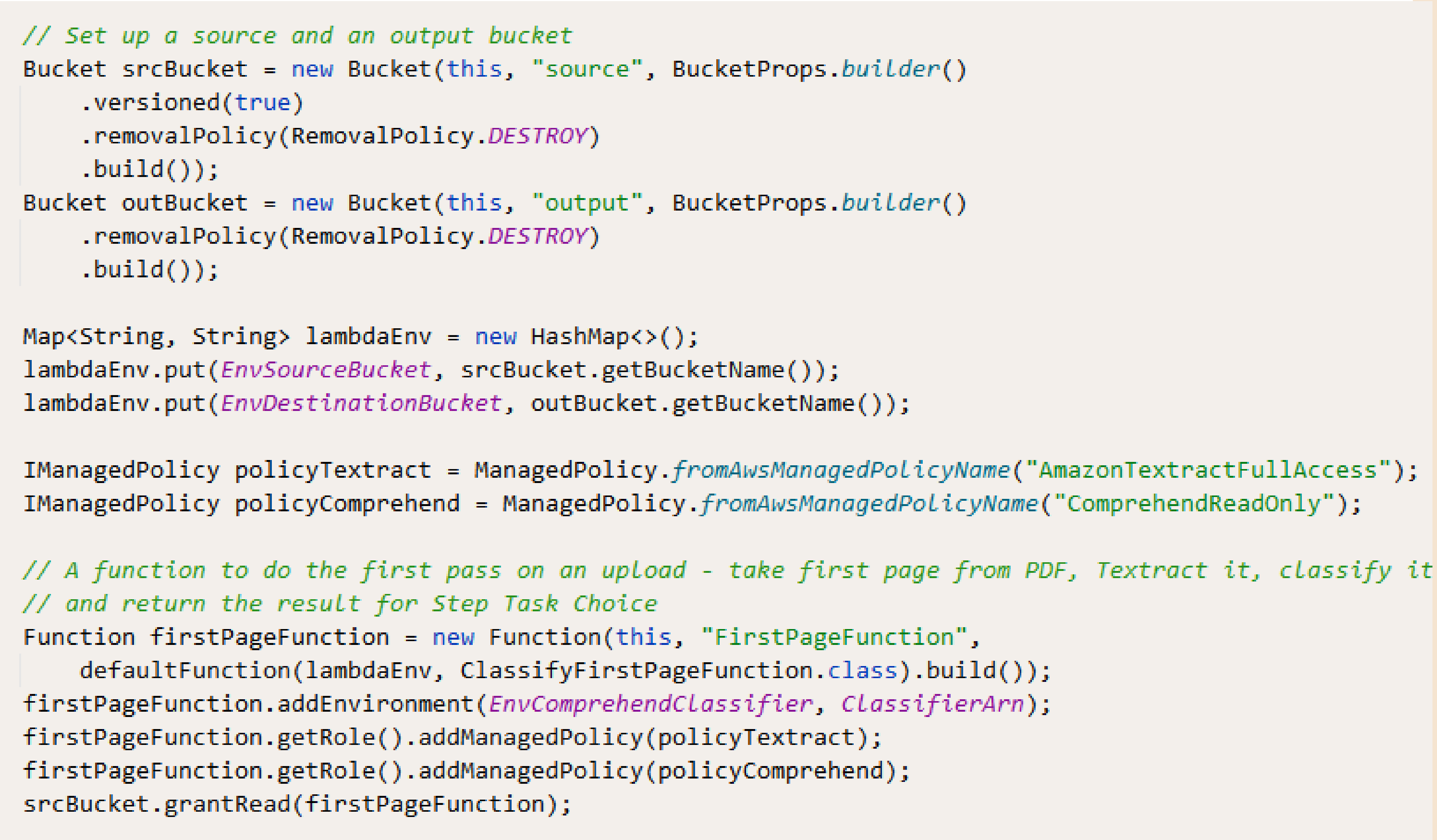


```
{
 "Resources": {
  "HelloCdkQueueB56C7789": {
   "Type": "AWS::SQS::Queue",
   "Properties": {
    "VisibilityTimeout": 300
   }
  },
  ...
```

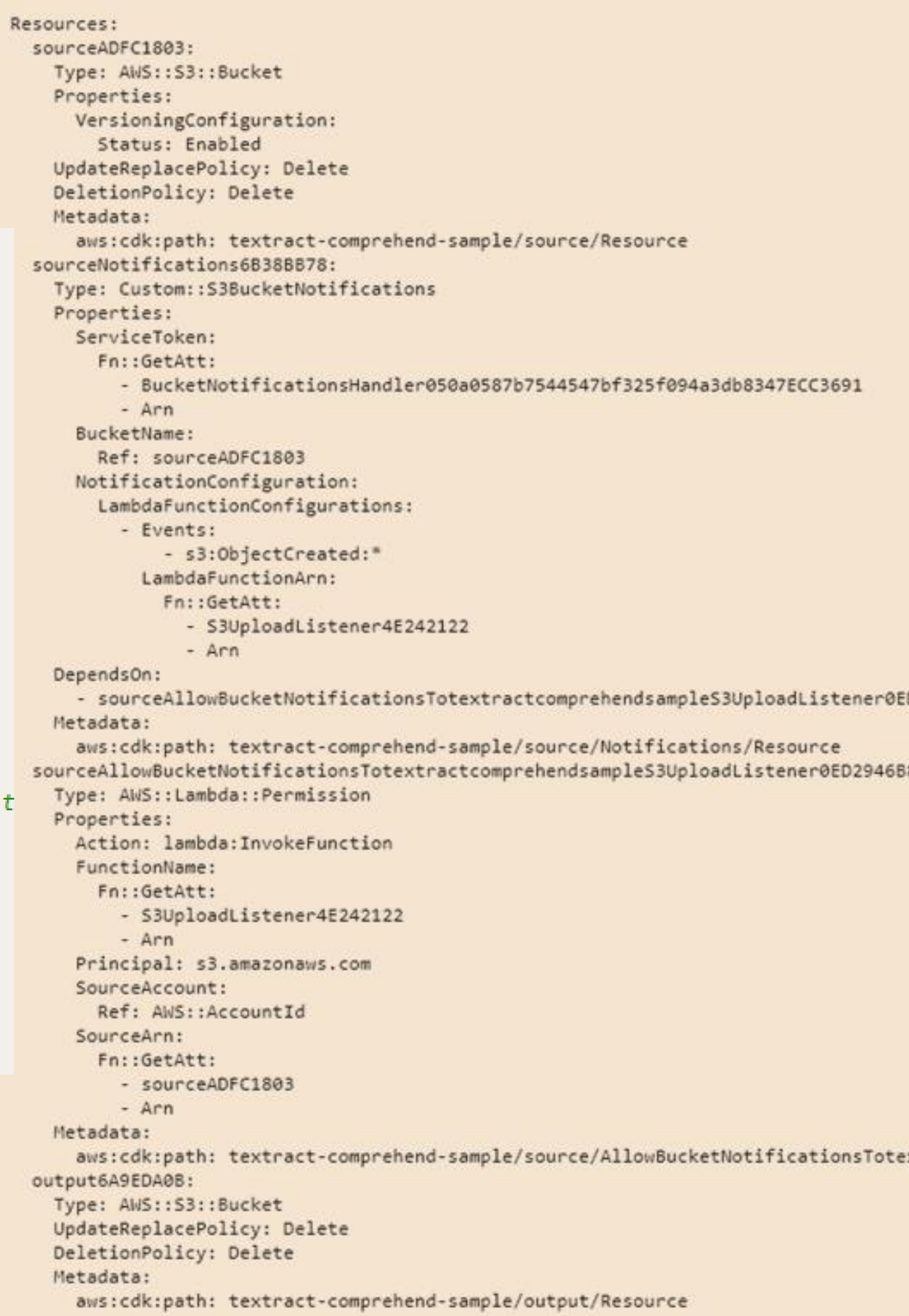


```
IAM Policy Changes
┌───┬──────────────────────────────────┬────────────────────────────────────────────────────────────┐
│   │ Resource                         │ Managed Policy ARN                                         │
├───┼──────────────────────────────────┼────────────────────────────────────────────────────────────┤
│ + │ ${FirstPageFunction/ServiceRole} │ arn:${AWS::Partition}:iam::aws:policy/ComprehendReadOnly   │
└───┴──────────────────────────────────┴────────────────────────────────────────────────────────────┘
(NOTE: There may be security-related changes not in this list. See https://github.com/aws/aws-cdk/issues/1299)

Do you wish to deploy these changes (y/n)? y
textract-comprehend-sample: deploying...
[0%] start: Publishing 7394859d2a36334877e81f52de0f7bbf7f30fe23cc9f41ff2a8e8095154fb79c:current
[100%] success: Published 7394859d2a36334877e81f52de0f7bbf7f30fe23cc9f41ff2a8e8095154fb79c:current
textract-comprehend-sample: creating CloudFormation changeset...
  0/12 | 2:47:42 pm | UPDATE_IN_PROGRESS   | AWS::CloudFormation::Stack         | textract-comprehend-sample User Initiated
 0/12 Currently in progress: textract-comprehend-sample
  0/12 | 2:48:22 pm | UPDATE_IN_PROGRESS   | AWS::IAM::Role                     | FirstPageFunction/ServiceRole (FirstPageFunctionServiceRole25AF35
A6)
  1/12 | 2:48:23 pm | UPDATE_IN_PROGRESS   | AWS::Lambda::Function              | StartTextractFunction (StartTextractFunctionCD0185FF)
  1/12 | 2:48:25 pm | UPDATE_COMPLETE      | AWS::Lambda::Function              | StartTextractFunction (StartTextractFunctionCD0185FF)
  2/12 | 2:48:40 pm | UPDATE_COMPLETE      | AWS::IAM::Role                     | FirstPageFunction/ServiceRole (FirstPageFunctionServiceRole25AF35
A6)
  3/12 | 2:48:44 pm | UPDATE_IN_PROGRESS   | AWS::Lambda::Function              | FirstPageFunction (FirstPageFunctionBEE9EA11)
  3/12 | 2:48:46 pm | UPDATE_COMPLETE      | AWS::Lambda::Function              | FirstPageFunction (FirstPageFunctionBEE9EA11)
  3/12 | 2:48:49 pm | UPDATE_IN_PROGRESS   | AWS::StepFunctions::StateMachine   | UploadFlow (UploadFlow6C932FD7)
  4/12 | 2:48:51 pm | UPDATE_COMPLETE      | AWS::StepFunctions::StateMachine   | UploadFlow (UploadFlow6C932FD7)
  7/12 | 2:48:55 pm | UPDATE_IN_PROGRESS   | AWS::Lambda::Function              | TextractCompletion (TextractCompletion5AE7AEDD)
  7/12 | 2:48:55 pm | UPDATE_IN_PROGRESS   | AWS::Lambda::Function              | S3UploadListener (S3UploadListener4E242122)
  7/12 | 2:48:56 pm | UPDATE_COMPLETE      | AWS::Lambda::Function              | TextractCompletion (TextractCompletion5AE7AEDD)
  7/12 | 2:48:56 pm | UPDATE_COMPLETE      | AWS::Lambda::Function              | S3UploadListener (S3UploadListener4E242122)
  7/12 | 2:48:58 pm | UPDATE_COMPLETE_CLEA | AWS::CloudFormation::Stack         | textract-comprehend-sample
  7/12 | 2:48:59 pm | UPDATE_COMPLETE      | AWS::CloudFormation::Stack         | textract-comprehend-sample

 ✅  textract-comprehend-sample

Stack ARN:
arn:aws:cloudformation:ap-southeast-2:[redacted]:stack/textract-comprehend-sample/3e68c070-5a09-11eb-9c86-06618561e0aa
```

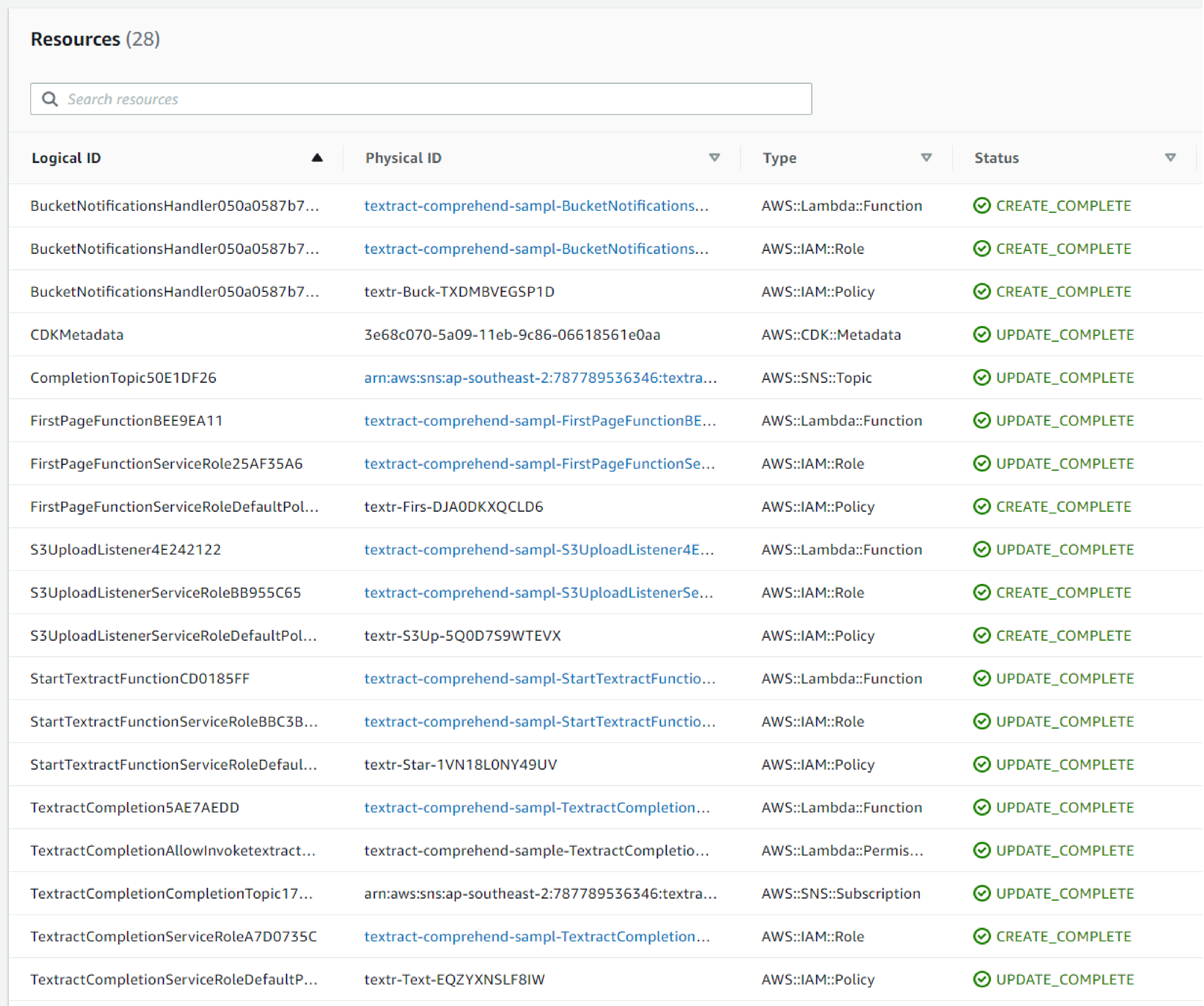
Creating resources using the CDK



```java
// Set up a source and an output bucket
Bucket srcBucket = new Bucket(this, "source", BucketProps.builder()
    .versioned(true)
    .removalPolicy(RemovalPolicy.DESTROY)
    .build());
Bucket outBucket = new Bucket(this, "output", BucketProps.builder()
    .removalPolicy(RemovalPolicy.DESTROY)
    .build());

Map<String, String> lambdaEnv = new HashMap<>();
lambdaEnv.put(EnvSourceBucket, srcBucket.getBucketName());
lambdaEnv.put(EnvDestinationBucket, outBucket.getBucketName());

IManagedPolicy policyTextract = ManagedPolicy.fromAwsManagedPolicyName("AmazonTextractFullAccess");
IManagedPolicy policyComprehend = ManagedPolicy.fromAwsManagedPolicyName("ComprehendReadOnly");

// A function to do the first pass on an upload - take first page from PDF, Textract it, classify it
// and return the result for Step Task Choice
Function firstPageFunction = new Function(this, "FirstPageFunction",
    defaultFunction(lambdaEnv, ClassifyFirstPageFunction.class).build());
firstPageFunction.addEnvironment(EnvComprehendClassifier, ClassifierArn);
firstPageFunction.getRole().addManagedPolicy(policyTextract);
firstPageFunction.getRole().addManagedPolicy(policyComprehend);
srcBucket.grantRead(firstPageFunction);
```



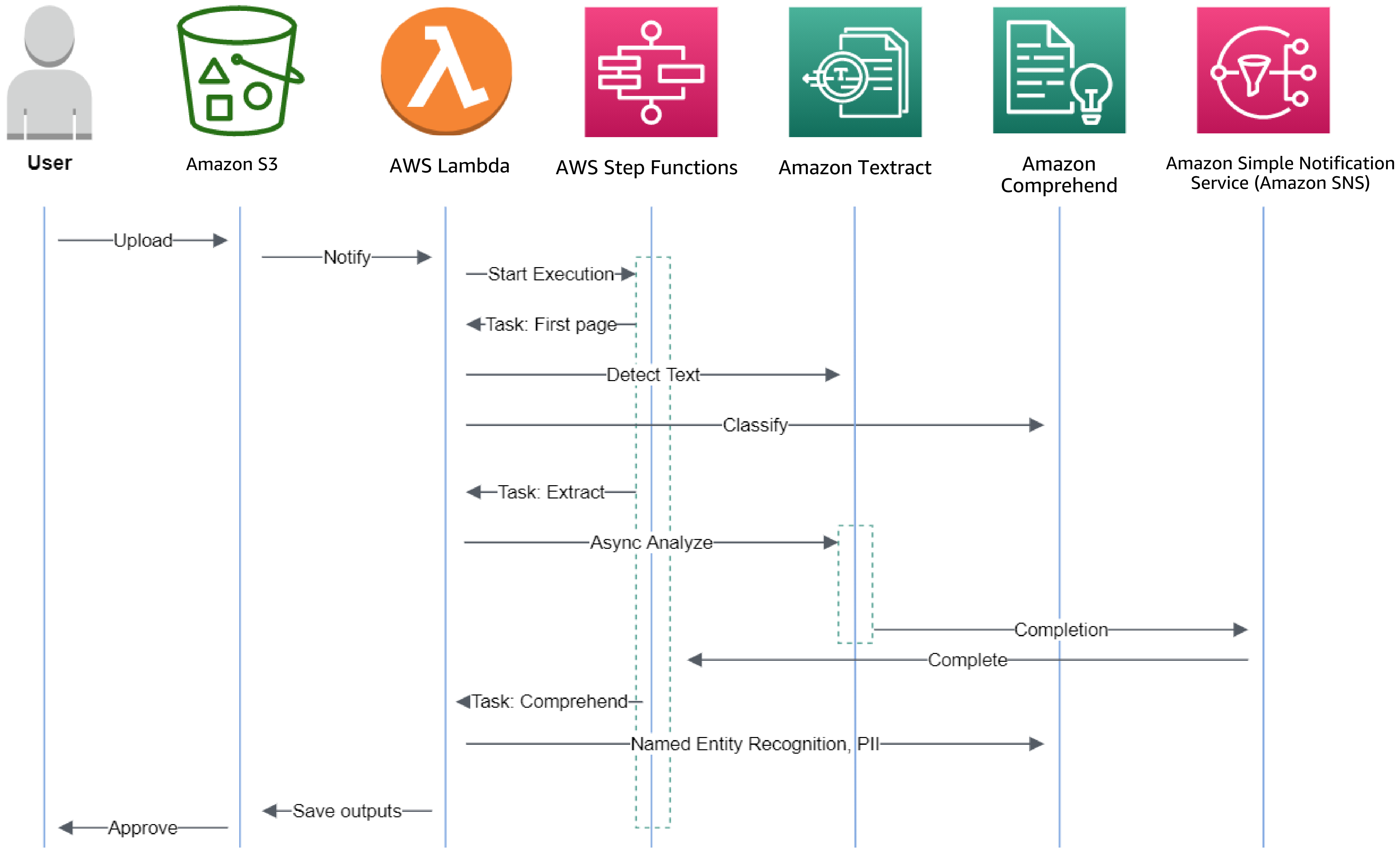
```yaml
Resources:
  sourceADFC1803:
    Type: AWS::S3::Bucket
    Properties:
      VersioningConfiguration:
        Status: Enabled
    UpdateReplacePolicy: Delete
    DeletionPolicy: Delete
    Metadata:
      aws:cdk:path: textract-comprehend-sample/source/Resource
  sourceNotifications6B38BB78:
    Type: Custom::S3BucketNotifications
    Properties:
      ServiceToken:
        Fn::GetAtt:
          - BucketNotificationsHandler050a0587b7544547bf325f094a3db8347ECC3691
          - Arn
      BucketName:
        Ref: sourceADFC1803
      NotificationConfiguration:
        LambdaFunctionConfigurations:
          - Events:
              - s3:ObjectCreated:*
            LambdaFunctionArn:
              Fn::GetAtt:
                - S3UploadListener4E242122
                - Arn
    DependsOn:
      - sourceAllowBucketNotificationsTotextractcomprehendsampleS3UploadListener0E
    Metadata:
      aws:cdk:path: textract-comprehend-sample/source/Notifications/Resource
  sourceAllowBucketNotificationsTotextractcomprehendsampleS3UploadListener0ED2946B
    Type: AWS::Lambda::Permission
    Properties:
      Action: lambda:InvokeFunction
      FunctionName:
        Fn::GetAtt:
          - S3UploadListener4E242122
          - Arn
      Principal: s3.amazonaws.com
      SourceAccount:
        Ref: AWS::AccountId
      SourceArn:
        Fn::GetAtt:
          - sourceADFC1803
          - Arn
    Metadata:
      aws:cdk:path: textract-comprehend-sample/source/AllowBucketNotificationsTote
  output6A9EDA0B:
    Type: AWS::S3::Bucket
    UpdateReplacePolicy: Delete
    DeletionPolicy: Delete
    Metadata:
      aws:cdk:path: textract-comprehend-sample/output/Resource
```



Resources (28)

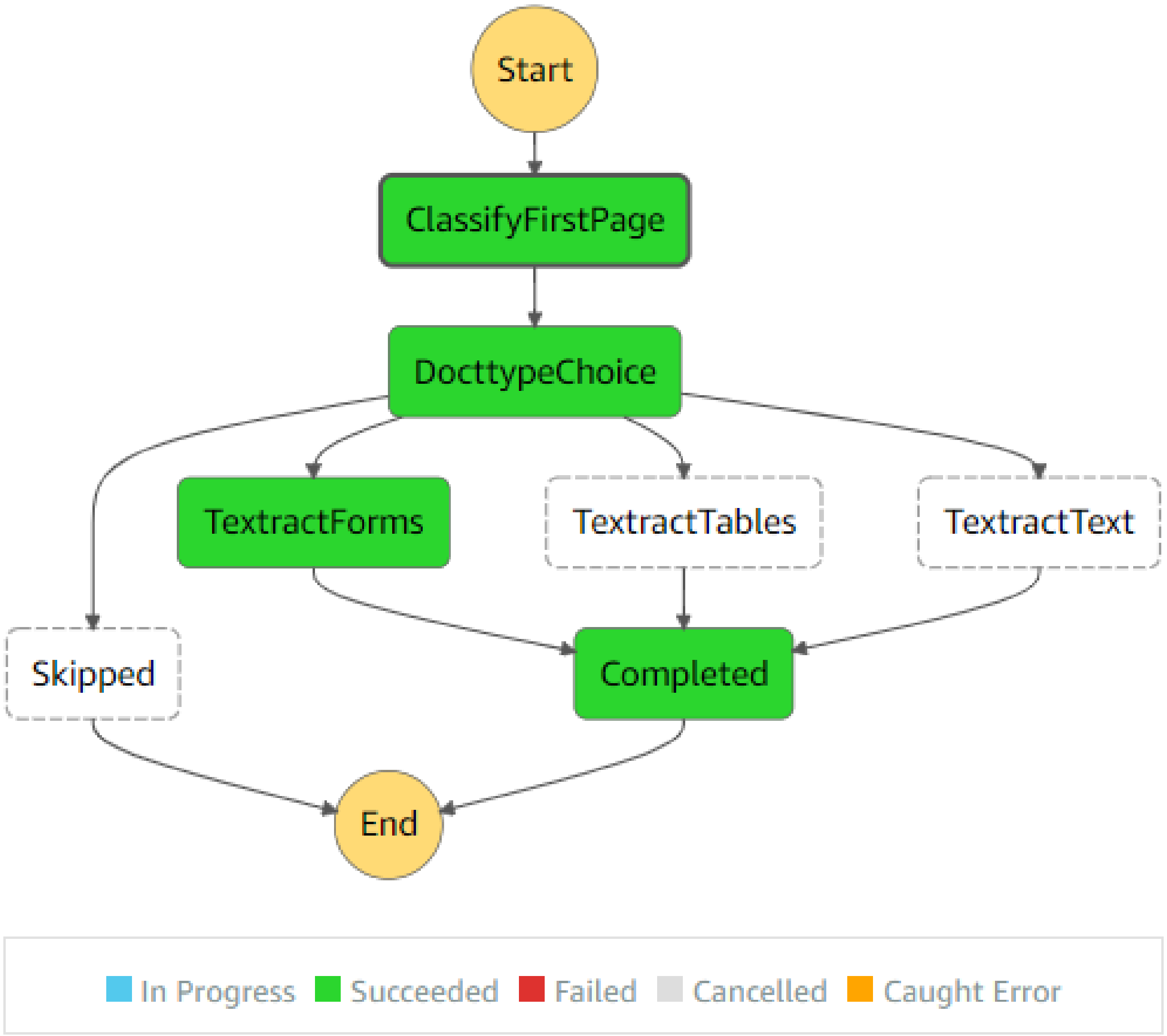
Logical ID	Physical ID	Type	Status
BucketNotificationsHandler050a0587b7…	textract-comprehend-sampl-BucketNotifications…	AWS::Lambda::Function	✓ CREATE_COMPLETE
BucketNotificationsHandler050a0587b7…	textract-comprehend-sampl-BucketNotifications…	AWS::IAM::Role	✓ CREATE_COMPLETE
BucketNotificationsHandler050a0587b7…	textr-Buck-TXDMBVEGSP1D	AWS::IAM::Policy	✓ CREATE_COMPLETE
CDKMetadata	3e68c070-5a09-11eb-9c86-06618561e0aa	AWS::CDK::Metadata	✓ UPDATE_COMPLETE
CompletionTopic50E1DF26	arn:aws:sns:ap-southeast-2:787789536346:textra…	AWS::SNS::Topic	✓ UPDATE_COMPLETE
FirstPageFunctionBEE9EA11	textract-comprehend-sampl-FirstPageFunctionBE…	AWS::Lambda::Function	✓ UPDATE_COMPLETE
FirstPageFunctionServiceRole25AF35A6	textract-comprehend-sampl-FirstPageFunctionSe…	AWS::IAM::Role	✓ UPDATE_COMPLETE
FirstPageFunctionServiceRoleDefaultPol…	textr-Firs-DJA0DKXQCLD6	AWS::IAM::Policy	✓ CREATE_COMPLETE
S3UploadListener4E242122	textract-comprehend-sampl-S3UploadListener4E…	AWS::Lambda::Function	✓ UPDATE_COMPLETE
S3UploadListenerServiceRoleBB955C65	textract-comprehend-sampl-S3UploadListenerSe…	AWS::IAM::Role	✓ CREATE_COMPLETE
S3UploadListenerServiceRoleDefaultPol…	textr-S3Up-5Q0D7S9WTEVX	AWS::IAM::Policy	✓ CREATE_COMPLETE
StartTextractFunctionCD0185FF	textract-comprehend-sampl-StartTextractFunctio…	AWS::Lambda::Function	✓ UPDATE_COMPLETE
StartTextractFunctionServiceRoleBBC3B…	textract-comprehend-sampl-StartTextractFunctio…	AWS::IAM::Role	✓ UPDATE_COMPLETE
StartTextractFunctionServiceRoleDefaul…	textr-Star-1VN18L0NY49UV	AWS::IAM::Policy	✓ UPDATE_COMPLETE
TextractCompletion5AE7AEDD	textract-comprehend-sampl-TextractCompletion…	AWS::Lambda::Function	✓ UPDATE_COMPLETE
TextractCompletionAllowInvoketextract…	textract-comprehend-sample-TextractCompletio…	AWS::Lambda::Permis…	✓ UPDATE_COMPLETE
TextractCompletionCompletionTopic17…	arn:aws:sns:ap-southeast-2:787789536346:textra…	AWS::SNS::Subscription	✓ UPDATE_COMPLETE
TextractCompletionServiceRoleA7D0735C	textract-comprehend-sampl-TextractCompletion…	AWS::IAM::Role	✓ CREATE_COMPLETE
TextractCompletionServiceRoleDefaultP…	textr-Text-EQZYXNSLF8IW	AWS::IAM::Policy	✓ UPDATE_COMPLETE

Sequence


User
Amazon S3
AWS Lambda
AWS Step Functions
Amazon Textract
Amazon Comprehend
Amazon Simple Notification Service (Amazon SNS)
Upload
Notify
Start Execution
Task: First page
Detect Text
Classify
Task: Extract
Async Analyze
Completion
Complete
Task: Comprehend
Named Entity Recognition, PII
Save outputs
Approve

aws INNOVATE
AI/ML EDITION

Demo



Start

ClassifyFirstPage

DocttypeChoice

TextractForms

TextractTables

TextractText

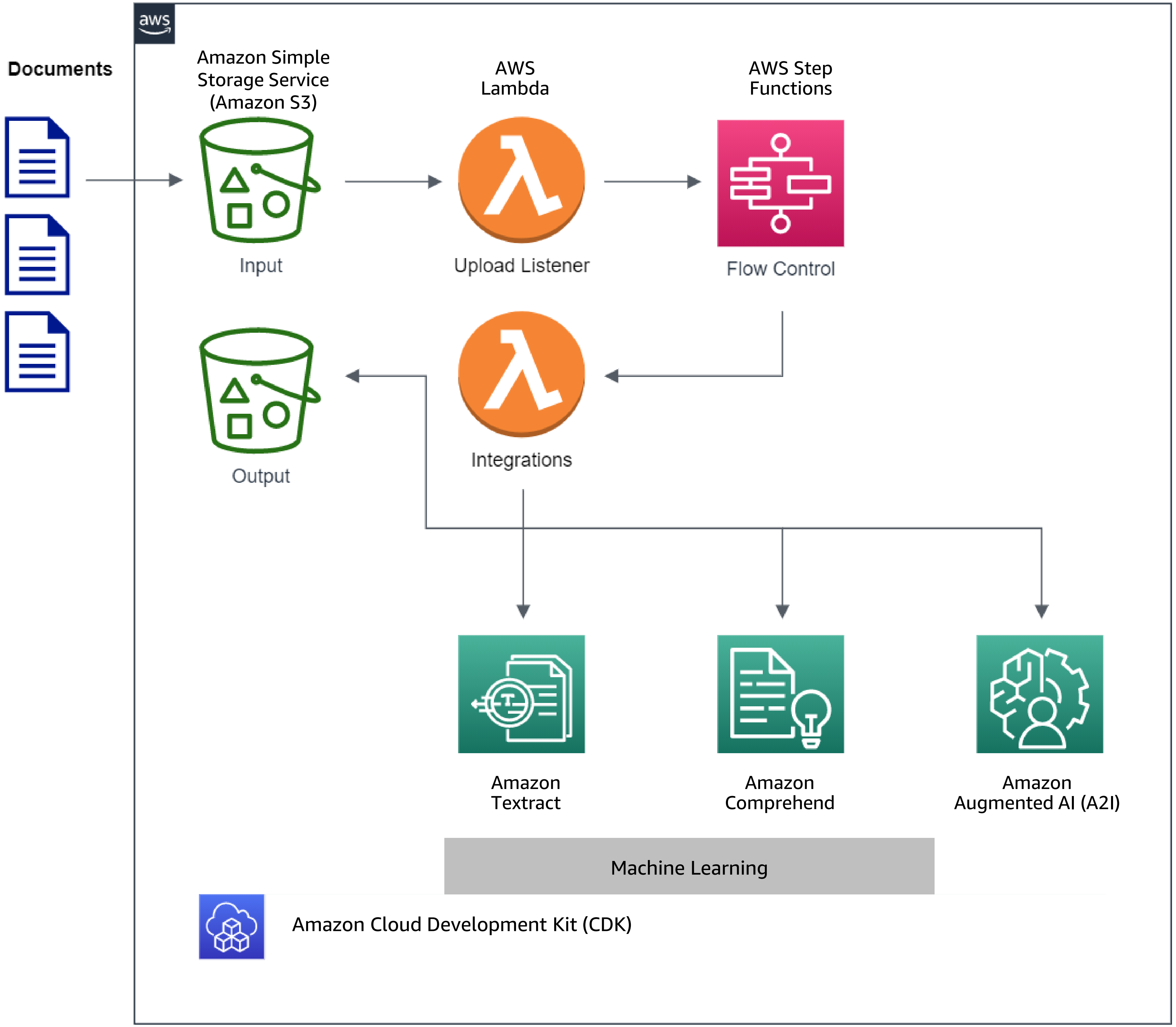
Skipped

Completed

End

🟦 In Progress 🟩 Succeeded 🟥 Failed ⬜ Cancelled 🟧 Caught Error

Architecture



Visit the AI and Machine Learning Resource Hub for more resources

Dive deeper with these resources, get inspired and learn how you can use machine learning to accelerate business outcomes.

- The machine learning journey e-book
- Machine learning enterprise guide
- 7 leading machine learning use cases e-book
- A strategic playbook for data, analytics, and machine learning
- Accelerating ML innovation through security e-book
- … and more!

https://tinyurl.com/aiml-aws

Visit resource hub »

AWS Machine Learning (ML) Training and Certification

Learn like an Amazonian, based on curriculum we’ve used to train our own developers and data scientists

AWS is how you build machine learning skills

Courses built on the curriculum leveraged by Amazon’s own teams. Learn from the experts at AWS.

Flexibility to Learn Your Way

Learn online with 65+ on-demand digital courses or live with virtual instructor-led training, plus hands-on labs and opportunities for practical application.

aws.training/machinelearning

Validate Your Expertise

Demonstrate expertise in building, training, tuning, and deploying machine learning models with an industry-recognized credential.

Thank You for Attending AWS Innovate

We hope you found it interesting! A kind reminder to **complete the survey.** Let us know what you thought of today's event and how we can improve the event experience for you in the future.

aws-apac-marketing@amazon.com

twitter.com/AWSCloud

facbook.com/AmazonWebServices

youtube.com/user/AmazonWebServices

slideshare.net/AmazonWebServices

twitch.tv/aws

aws INNOVATE
AI/ML EDITION

Thank you!