Process Documents Faster Using AWS Machine Learning

Sameer Karnik, PM – Comprehend, AWS

Gillian Armstrong, Solutions Engineer, Liberty Information Technology

Dave McFadden, Solutions Engineer, Liberty Information Technology

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

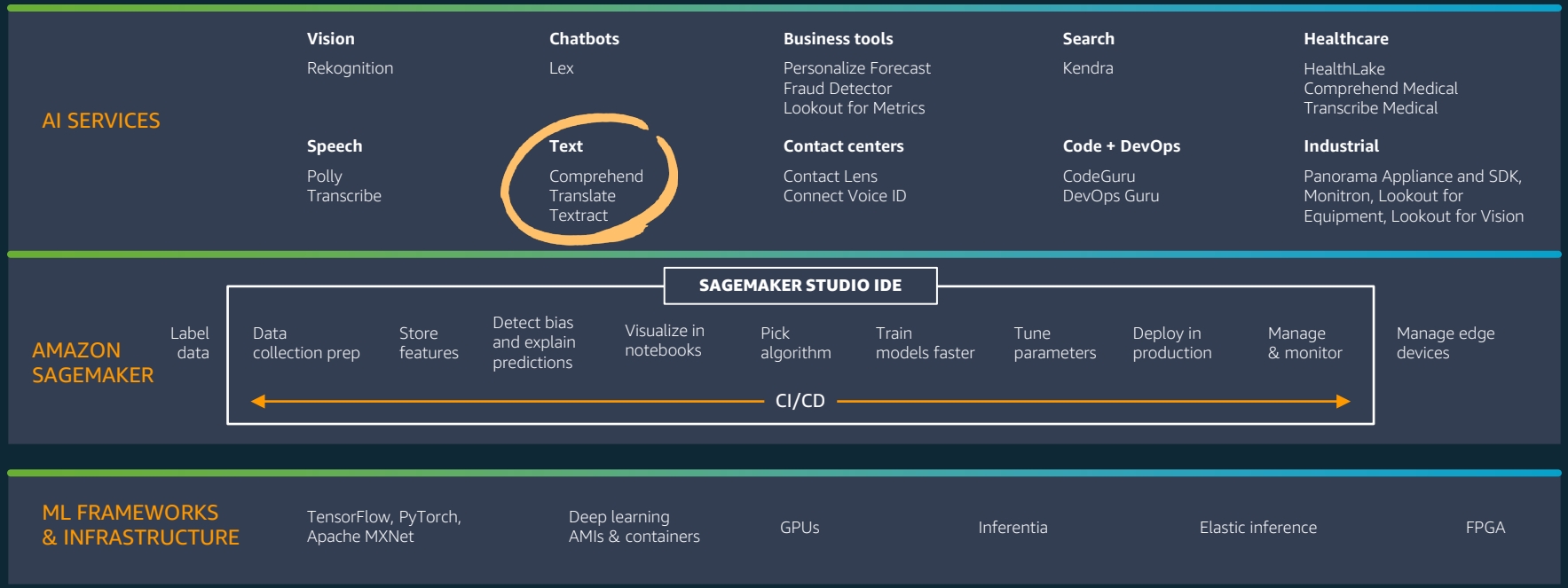
Our mission at AWS

Put machine learning (ML) in the hands of every developer

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The AWS ML stack

Broadest and most complete set of machine learning capabilities



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Intelligent Document Processing (IDP) use case

Billions of documents created globally everyday in physical and various digital formats with varying relevance to every reader

Experts trained to understand relevance read and extract information for people and processes to consume

Typically, automation applies rules-based or templates-based extraction that are narrowly focused on specific formats and use cases

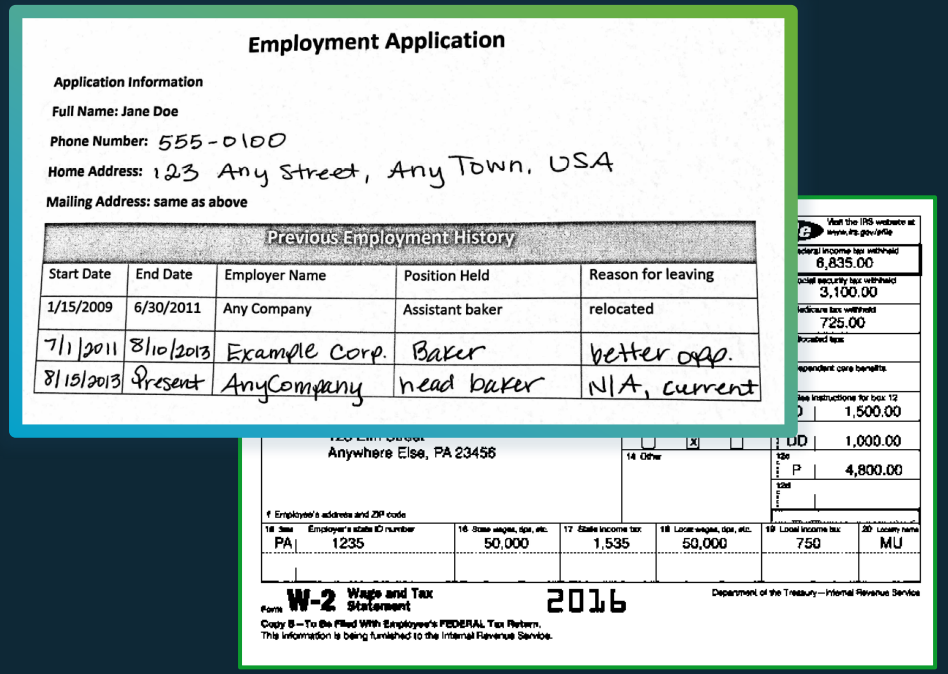
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Processing structured documents

Output

Full name	**First:** Jane **Middle:** X **Last:** Doe
Phone Number	**Number:** 555-0100
Home address	**Street number:** 123 **Street name:** Any Street **City:** Any town **Country:** USA
Mailing address	Same as above

- Forms/key-value pairs
- Tables
- Logical groupings
- Relationships
- Glyphs



Employment Application

Application Information

Full Name: Jane Doe

Phone Number: 555-0100

Home Address: 123 Any Street, Any Town, USA

Mailing Address: same as above

Previous Employment History

Start Date	End Date	Employer Name	Position Held	Reason for leaving
1/15/2009	6/30/2011	Any Company	Assistant baker	relocated
7/1/2011	8/10/2013	Example Corp.	Baker	better opp.
8/15/2013	Present	AnyCompany	head baker	N/A, current

Anywhere Else, PA 23456

16 State	Employer's state ID number	16 State wages, tips, etc.	17 State income tax	18 Local wages, tips, etc.	19 Local income tax	20 Locality name
PA	1235	50,000	1,535	50,000	750	MU

Form W-2 Wage and Tax Statement 2016

Department of the Treasury—Internal Revenue Service

Copy B—To Be Filed With Employee's FEDERAL Tax Return.

This information is being furnished to the Internal Revenue Service.

6,835.00

3,100.00

725.00

1,500.00

DD 1,000.00

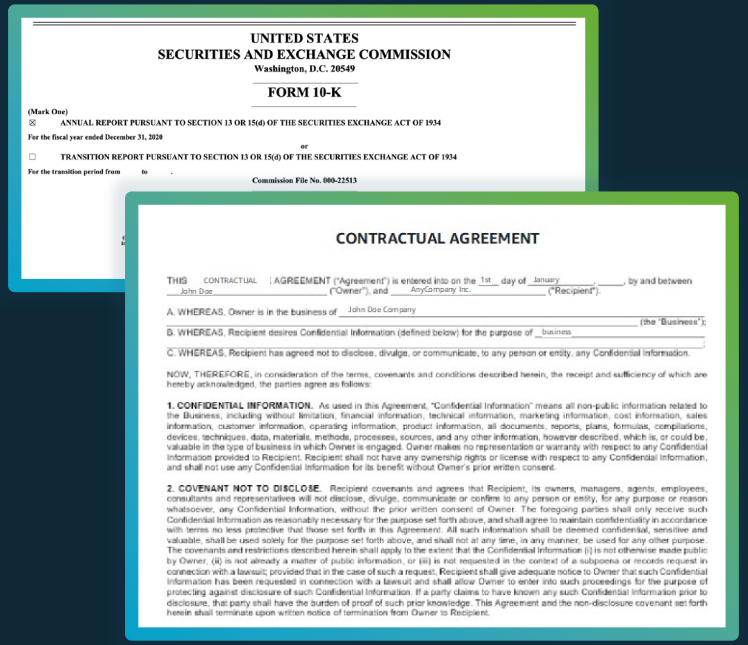
P 4,800.00

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Processing semi-structured and unstructured documents

- Context-aware, NLP-based, no regex
- Pre-trained entities extraction, e.g., person, dates, organizations, quantities, PII elements etc.
- Easy to train to extract custom entity types
- Customer-defined document classification labels



UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 10-K

(Mark One)
☒ ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934
For the fiscal year ended December 31, 2020
or
☐ TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934
For the transition period from to .
Commission File No. 000-22513

CONTRACTUAL AGREEMENT

THIS CONTRACTUAL AGREEMENT ("Agreement") is entered into on the 1st day of January, ____, by and between John Doe ("Owner"), and AnyCompany Inc. ("Recipient").

A. WHEREAS, Owner is in the business of John Doe Company (the "Business");

B. WHEREAS, Recipient desires Confidential Information (defined below) for the purpose of business;

C. WHEREAS, Recipient has agreed not to disclose, divulge, or communicate, to any person or entity, any Confidential Information.

NOW, THEREFORE, in consideration of the terms, covenants and conditions described herein, the receipt and sufficiency of which are hereby acknowledged, the parties agree as follows:

1. CONFIDENTIAL INFORMATION. As used in this Agreement, "Confidential Information" means all non-public information related to the Business, including without limitation, financial information, technical information, marketing information, cost information, sales information, customer information, operating information, product information, all documents, reports, plans, formulae, compilations, devices, techniques, data, materials, methods, processes, sources, and any other information, however described, which is, or could be, valuable in the type of business in which Owner is engaged. Owner makes no representation or warranty with respect to any Confidential Information provided to Recipient. Recipient shall not have any ownership rights or license with respect to any Confidential Information, and shall not use any Confidential Information for its benefit without Owner's prior written consent.

2. COVENANT NOT TO DISCLOSE. Recipient covenants and agrees that Recipient, its owners, managers, agents, employees, consultants and representatives will not disclose, divulge, communicate or confirm to any person or entity, for any purpose or reason whatsoever, any Confidential Information, without the prior written consent of Owner. The foregoing parties shall only receive such Confidential Information as reasonably necessary for the purpose set forth above, and shall agree to maintain confidentiality in accordance with terms no less protective that those set forth in this Agreement. All such information shall be deemed confidential, sensitive and valuable, shall be used solely for the purpose set forth above, and shall not at any time, in any manner, be used for any other purpose. The covenants and restrictions described herein shall apply to the extent that the Confidential Information (i) is not otherwise made public by Owner, (ii) is not already a matter of public information, or (iii) is not requested in the context of a subpoena or records request in connection with a lawsuit; provided that in the case of such a request, Recipient shall give adequate notice to Owner that such Confidential Information has been requested in connection with a lawsuit and shall allow Owner to enter into such proceedings for the purpose of protecting against disclosure of such Confidential Information. If a party claims to have known any such Confidential Information prior to disclosure, that party shall have the burden of proof of such prior knowledge. This Agreement and the non-disclosure covenant set forth herein shall terminate upon written notice of termination from Owner to Recipient.

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Intelligent document processing with Amazon AI Services

Amazon Textract

Extract text and data from virtually any document

Amazon Comprehend

A fully-managed and continuously trained service that discovers insights and relationships in text

Amazon Augmented AI

Easily implement human review of machine learning predictions

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

Document Extraction at Liberty Mutual

Gillian Armstrong
Dave McFadden

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

The Challenge

Across our entire Value Chain lots of transactions are performed via “paper” (PDF, Text Documents, Spreadsheets, Images, Scans) sent from our customers.

Pricing and Underwriting → Sales and Distribution → Servicing and Policy Admin → Claims Management

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The Challenge

Our employees spend millions of hours manually inputting this data into our business systems.

Because of the high effort, a lot of data is never connected to related data in the organization.

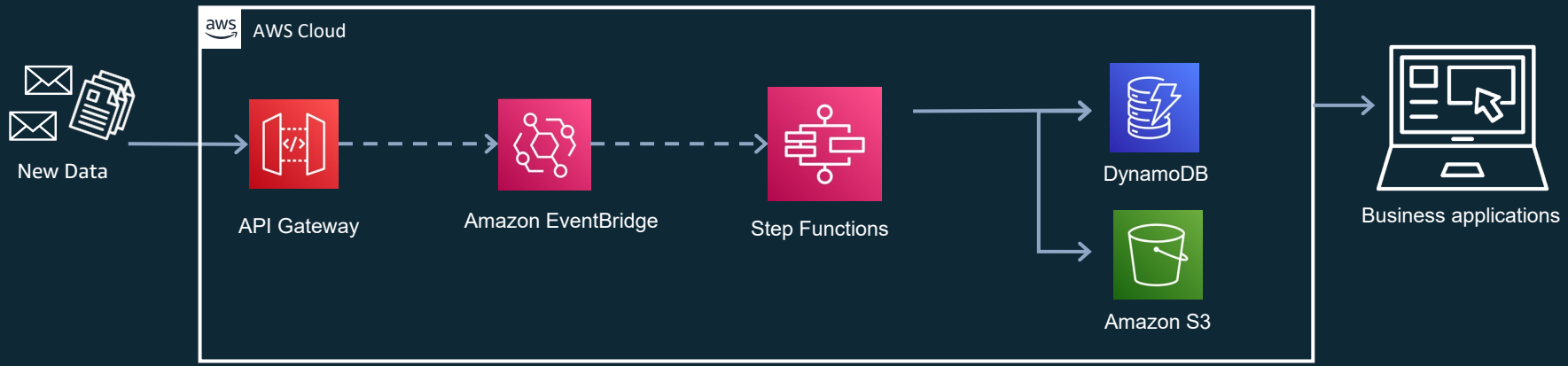
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The Opportunity

Find ways to improve efficiencies across the entire value chain through AI-augmented Automation

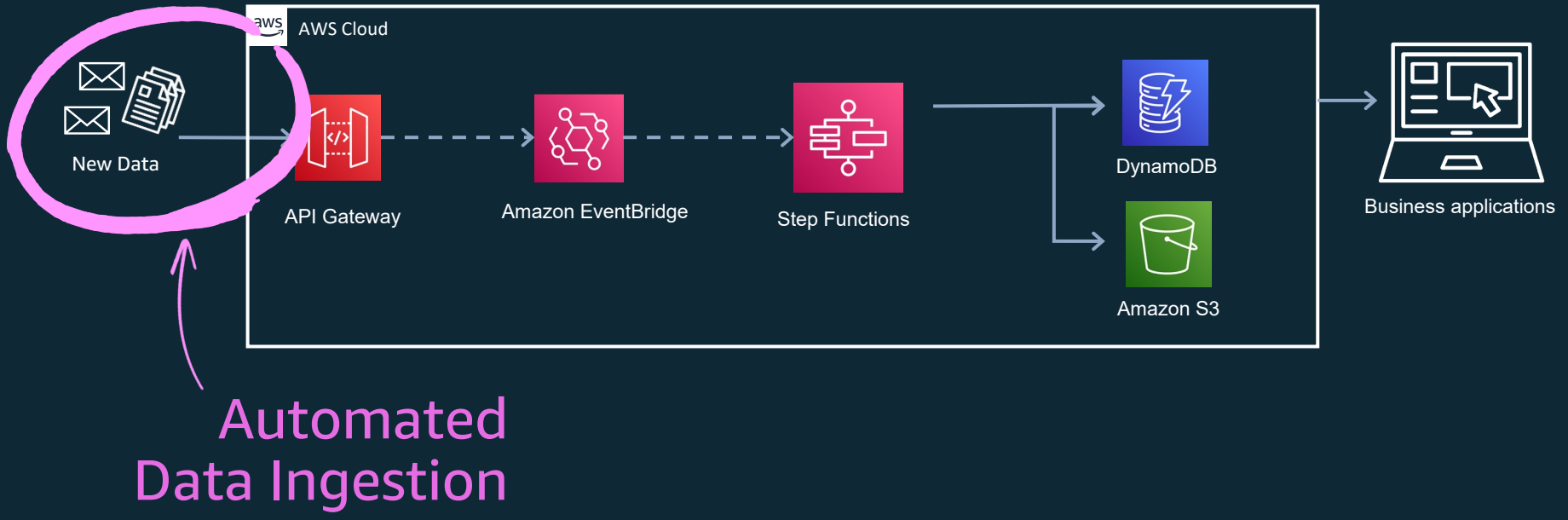
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



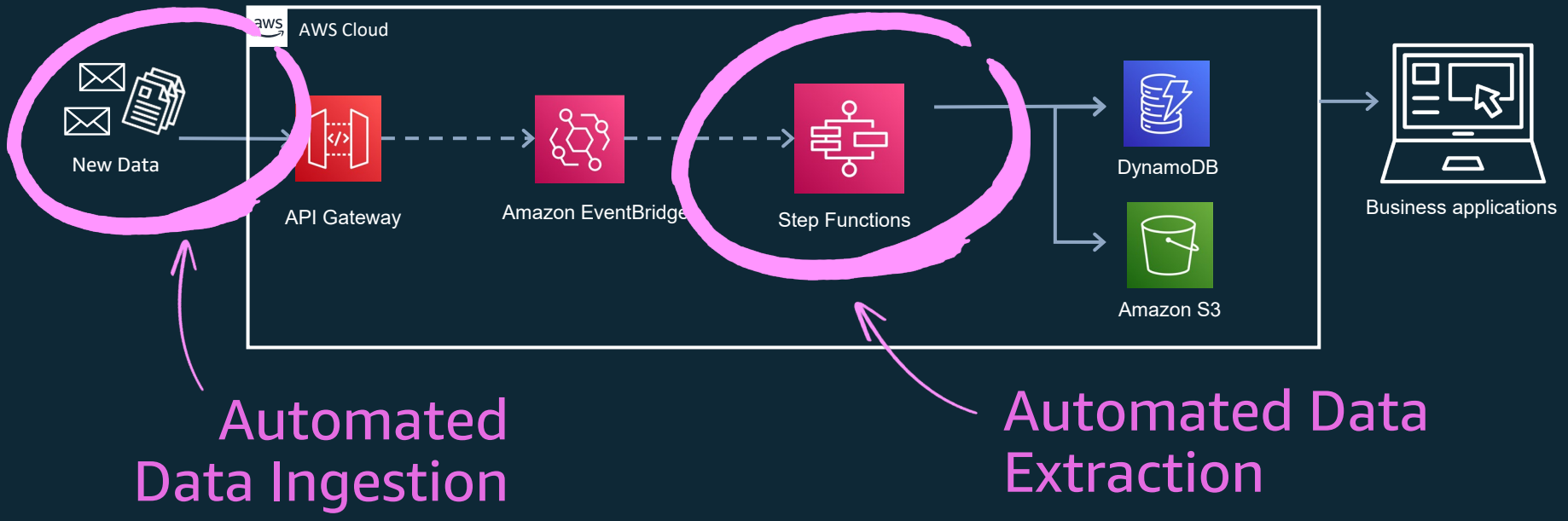
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



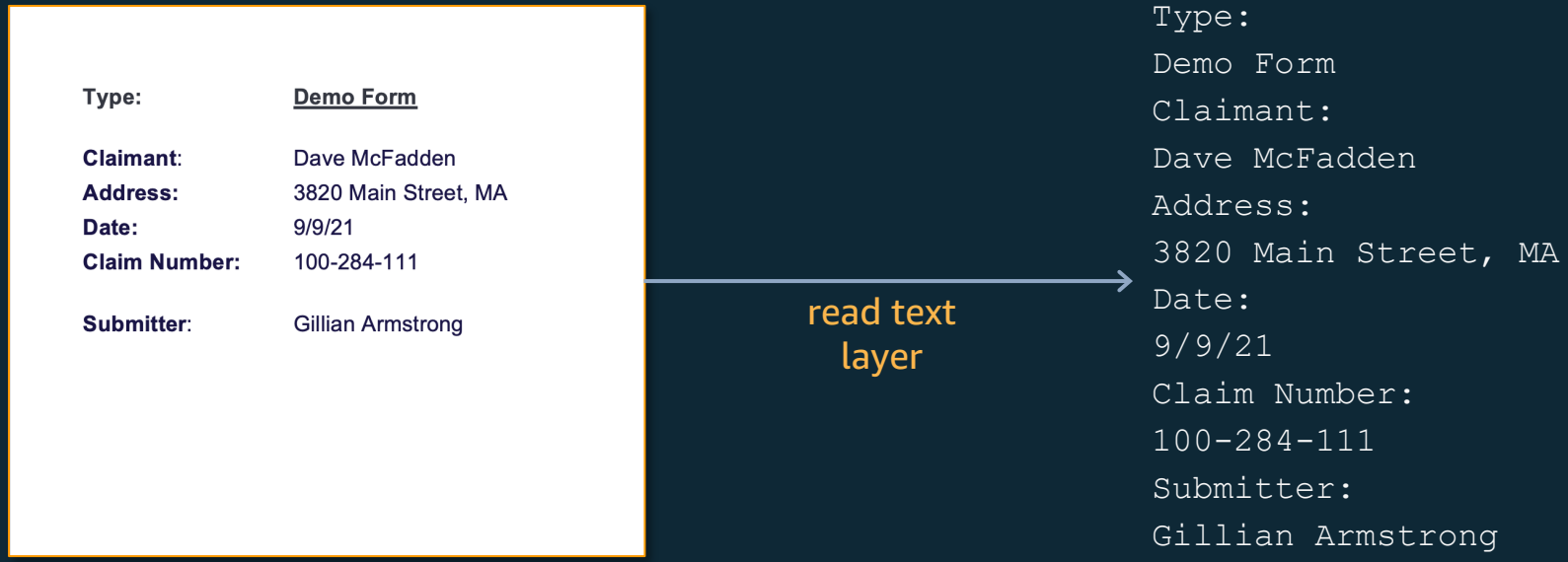
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated Data Extraction

Data Extraction doesn't just save the data (e.g., PDF form), it saves the Information contained in that form

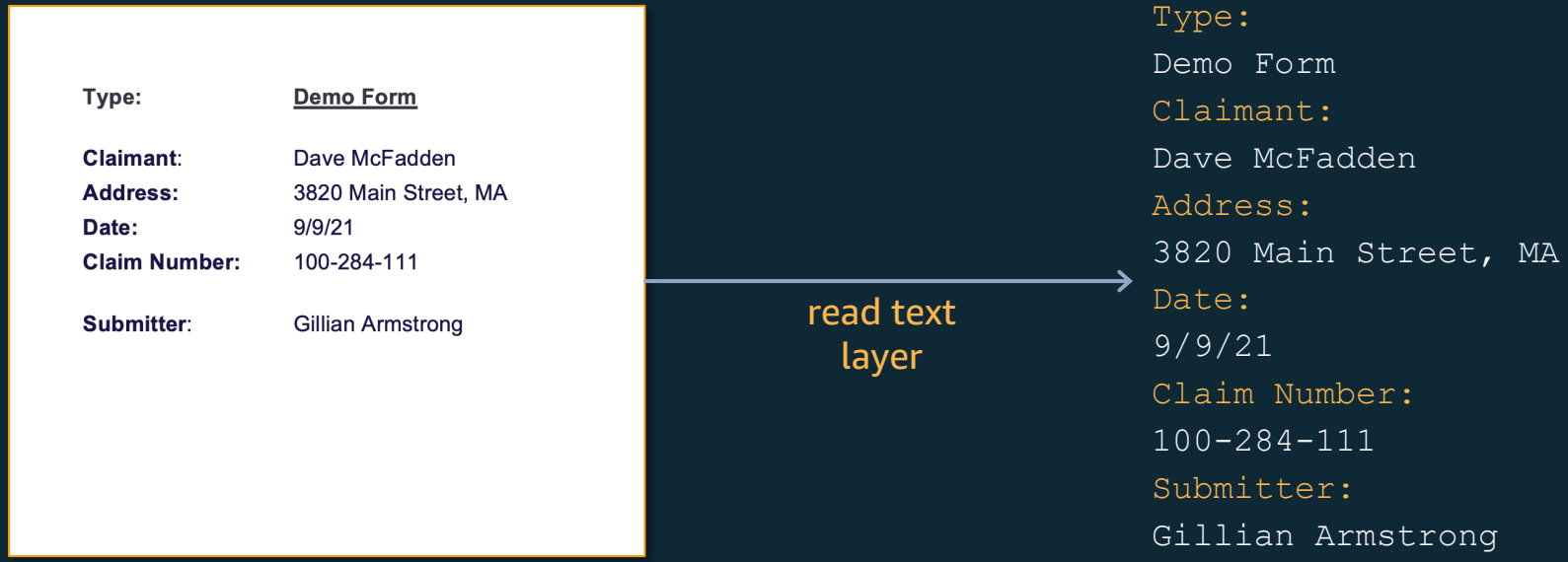
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Machine Readable Text



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Machine Readable Text



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

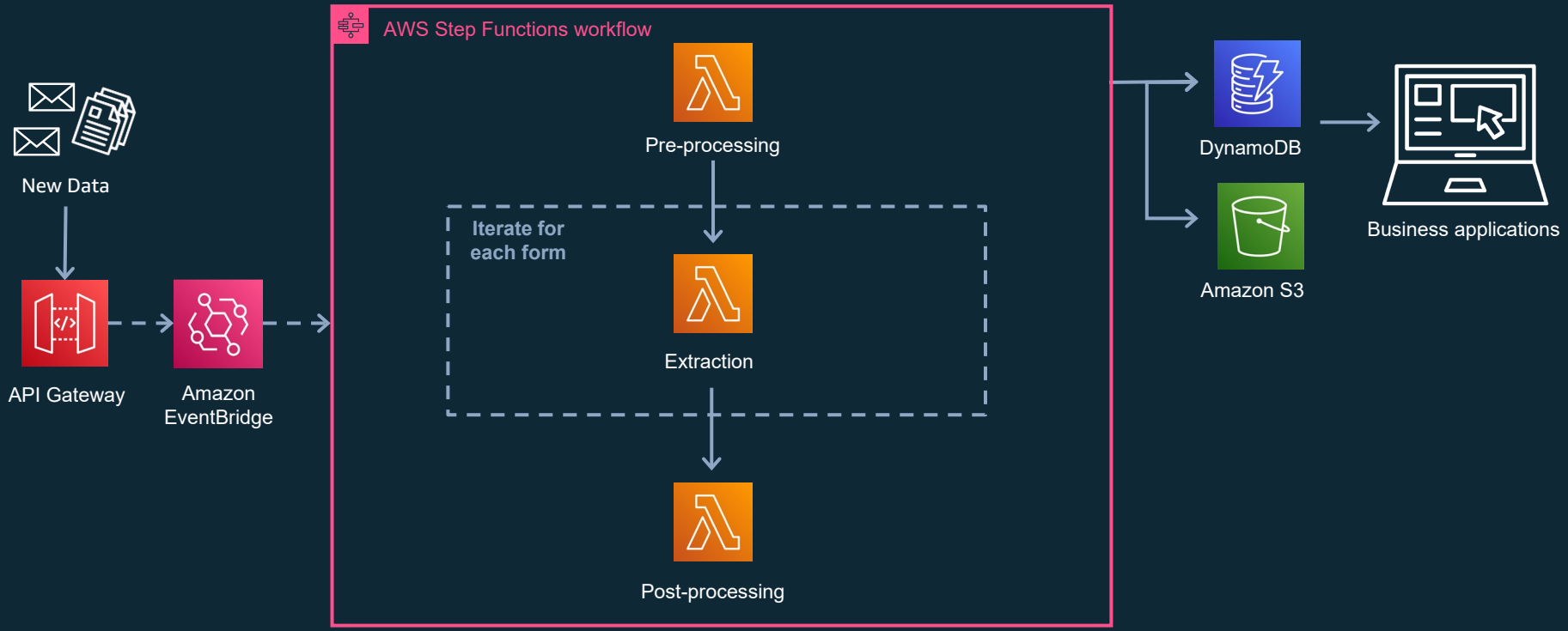
Automated Data Extraction

We need to take that Text block and transform it into Structured Tabular Data

Type	Claimant	Address	Date	Claim Number	Submitter
Demo Form	Dave McFadden	3820 Main Street, MA	9/9/21	100-284-111	Gillian Armstrong

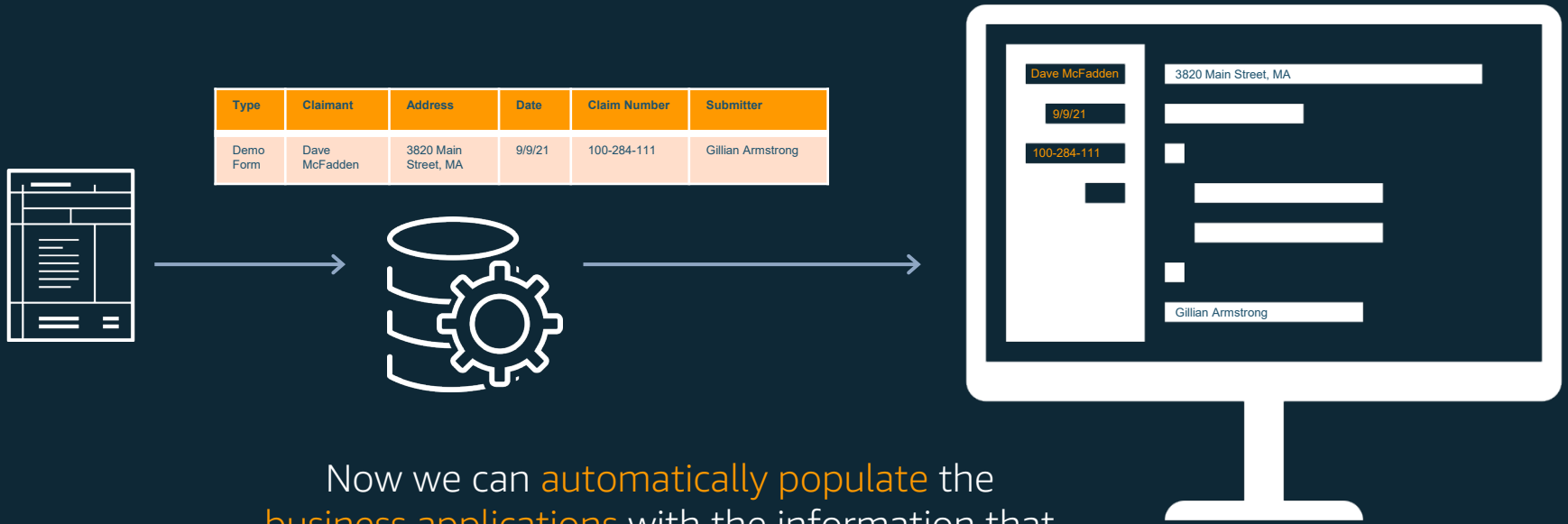
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

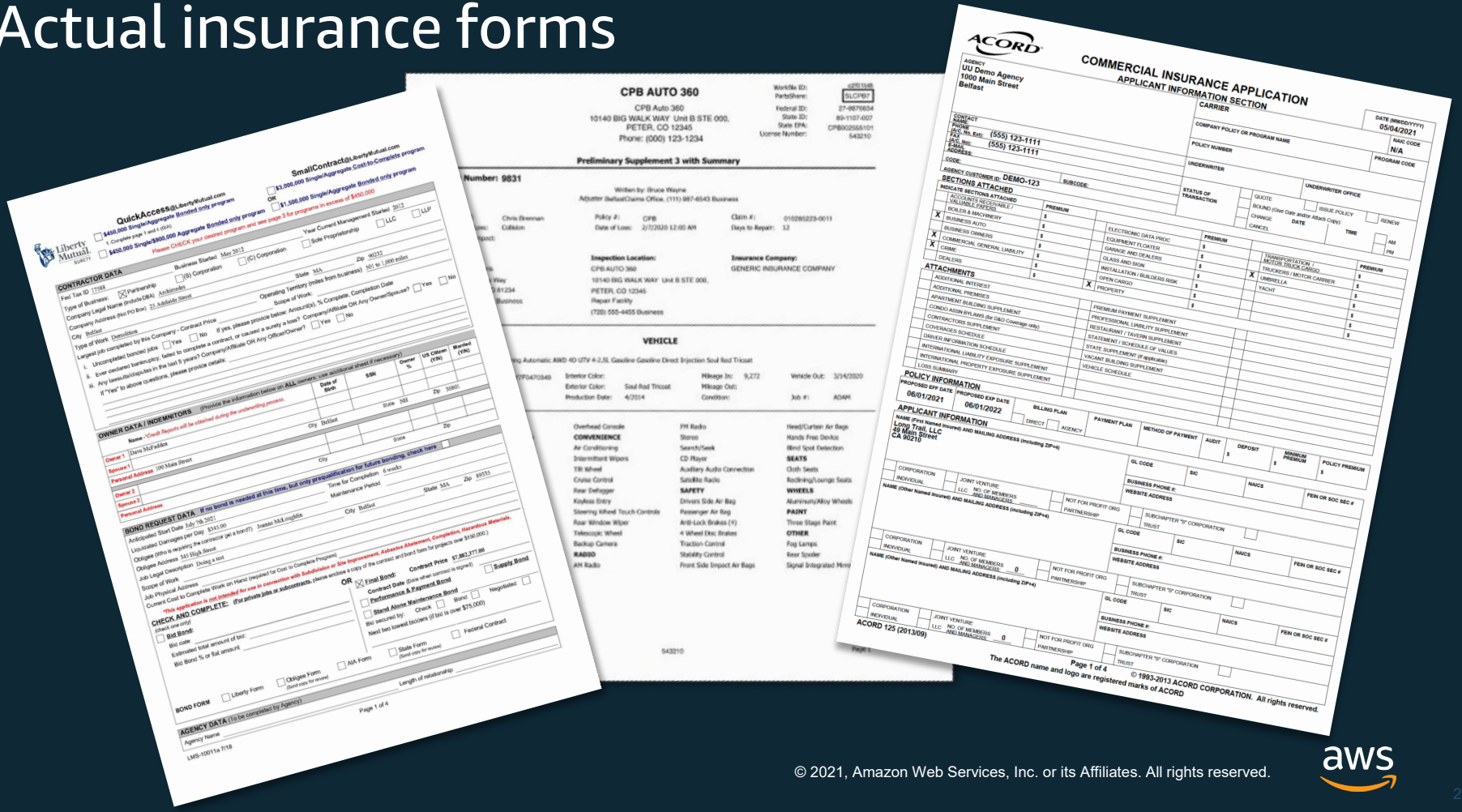
Automated Data Extraction



Now we can automatically populate the business applications with the information that previously would have been manually input

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Actual insurance forms



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Machine Readable Text

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Machine Readable Text

read text layer

```
Type: Demo Form
Claimant: Dave McFadden
Address
3820 Main Street, MA
Date:
Claim Number:
9/9/21
100-284-111
Gillian Armstrong
Submitter
```

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

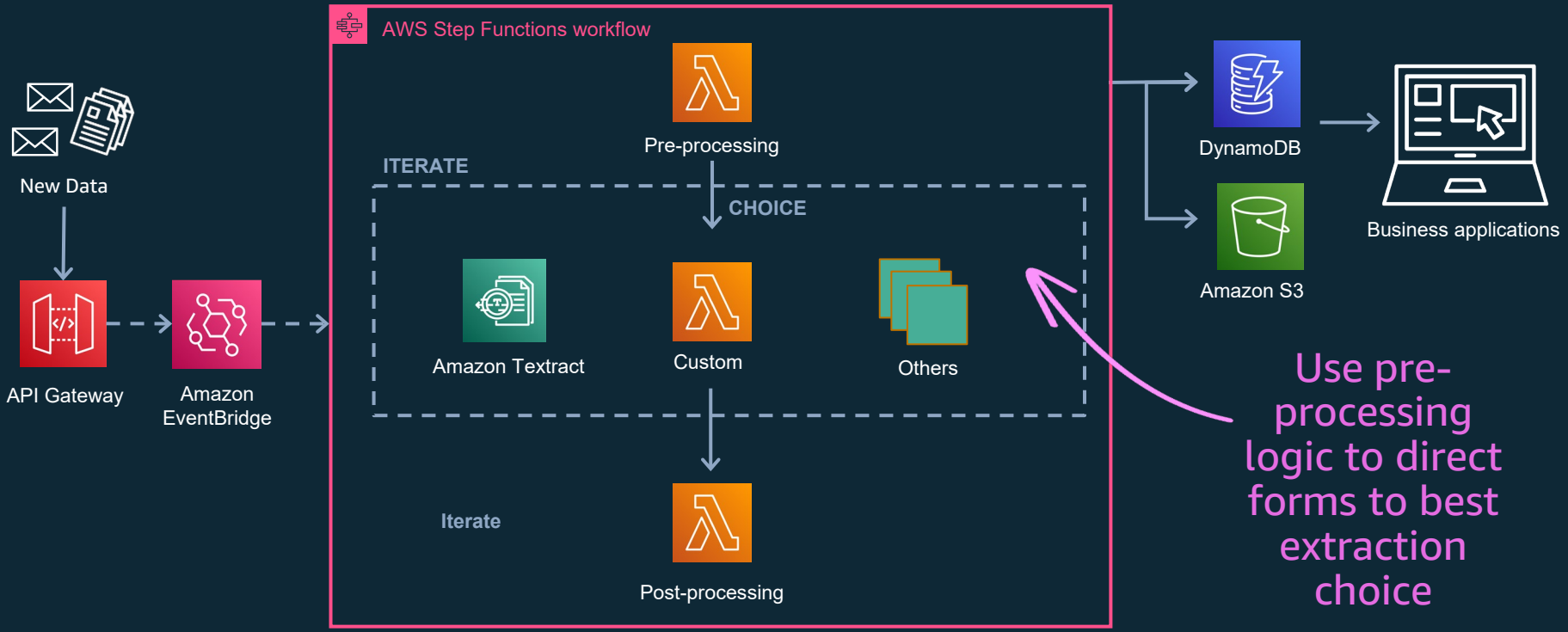
Machine Readable Text

read text layer

```
Type: Demo Form
Claimant: Dave McFadden
Address
3820 Main Street, MA
Date:
Claim Number:
9/9/21
100-284-111
Gillian Armstrong
Submitter
```

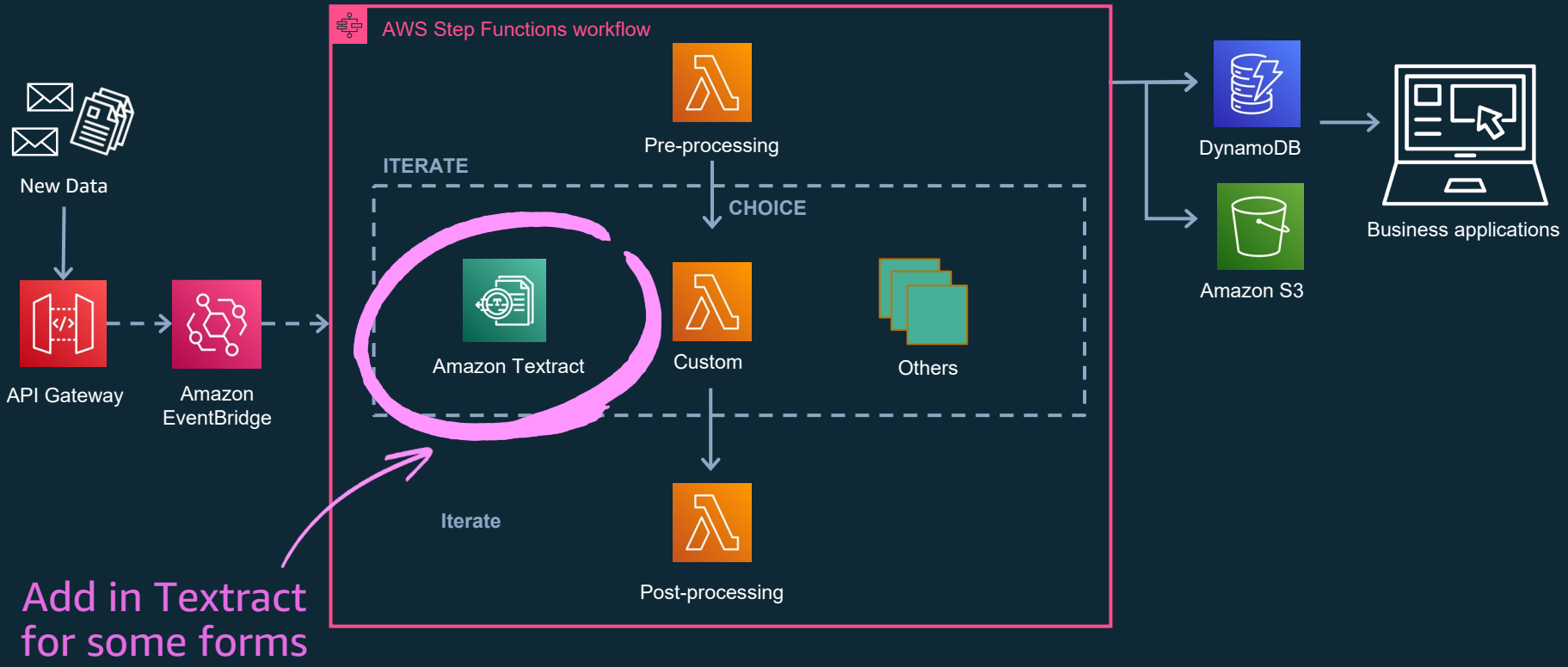
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

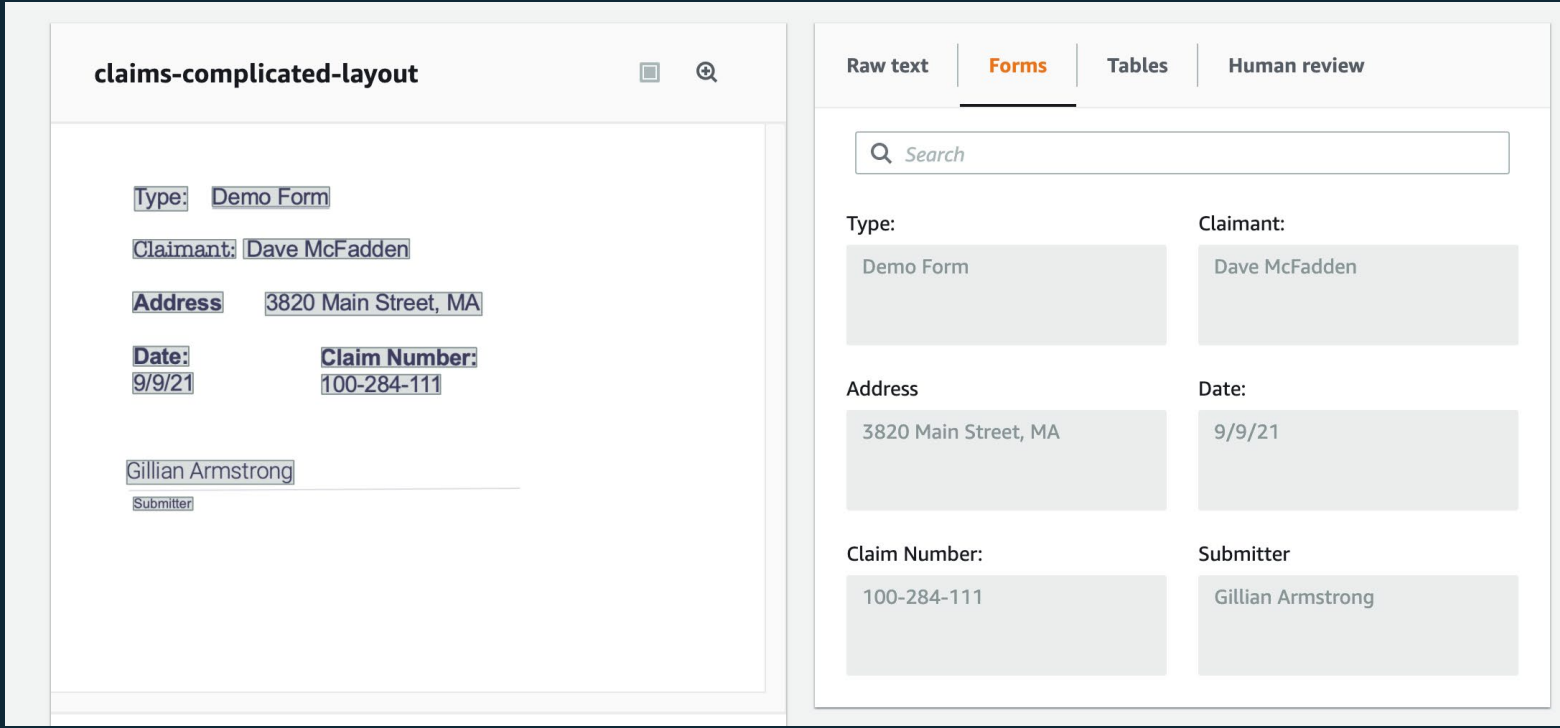
Machine Readable Text

read text layer

```
Type: Demo Form
Claimant: Dave McFadden
Address
3820 Main Street, MA
Date:
Claim Number:
9/9/21
100-284-111
Gillian Armstrong
Submitter
```

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Textract forms API



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

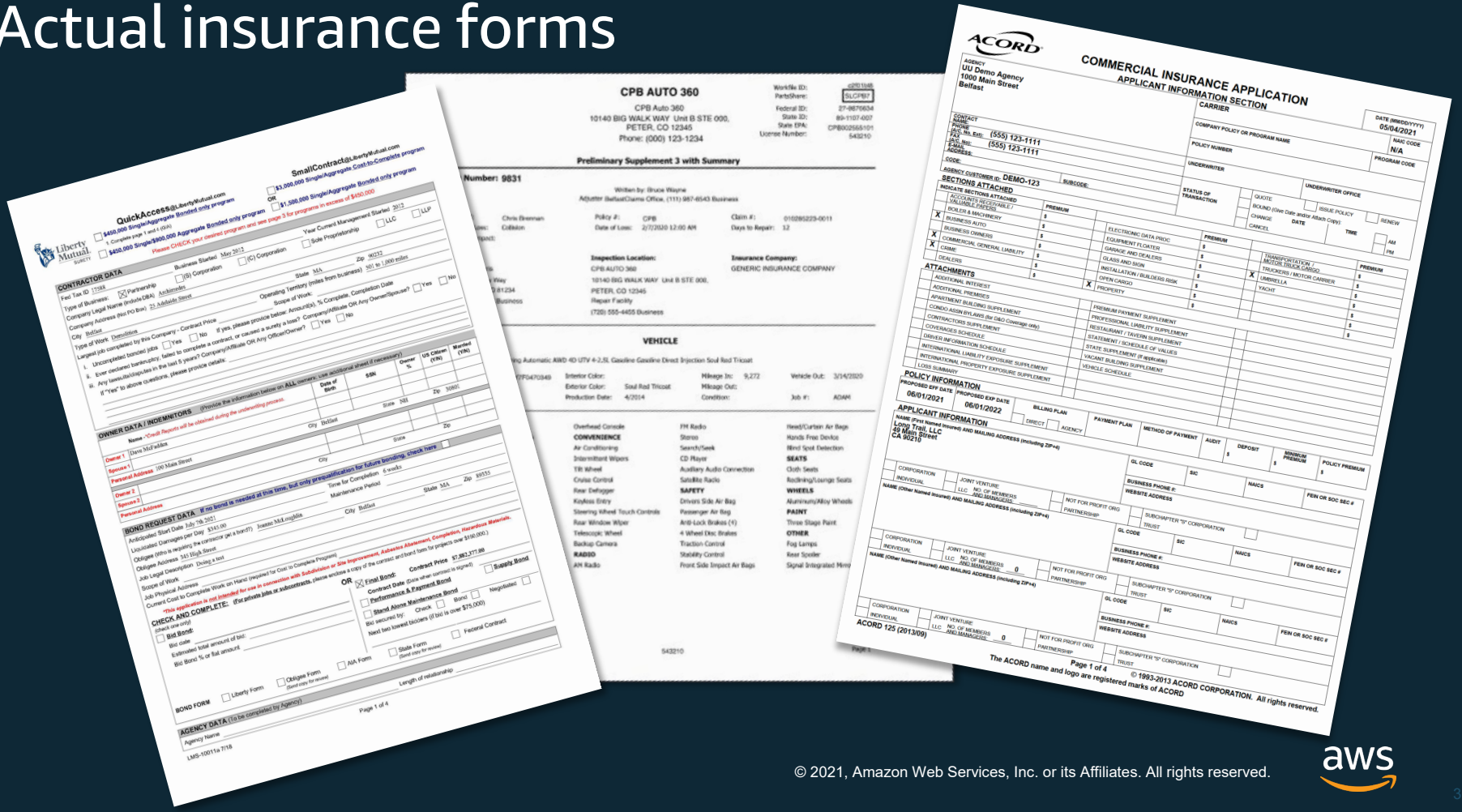
Textract forms service

read key-value pairs

Key	Value
Type:	Demo Form
Claimant:	Dave McFadden
Address	3820 Main Street, MA
Date:	09/09/2021
Claim Number:	100-284-111
Submitter	Gillian Armstrong

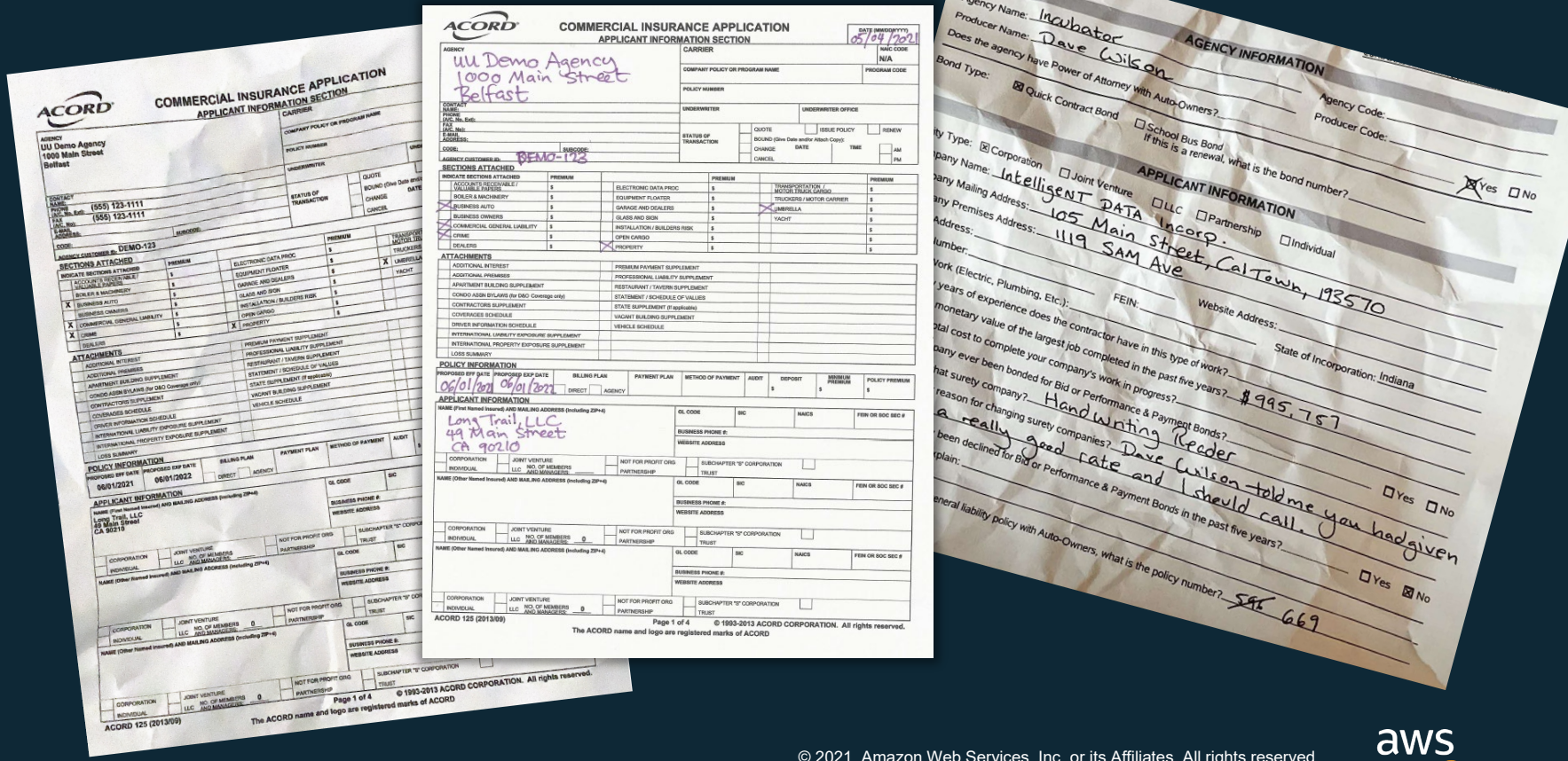
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Actual insurance forms



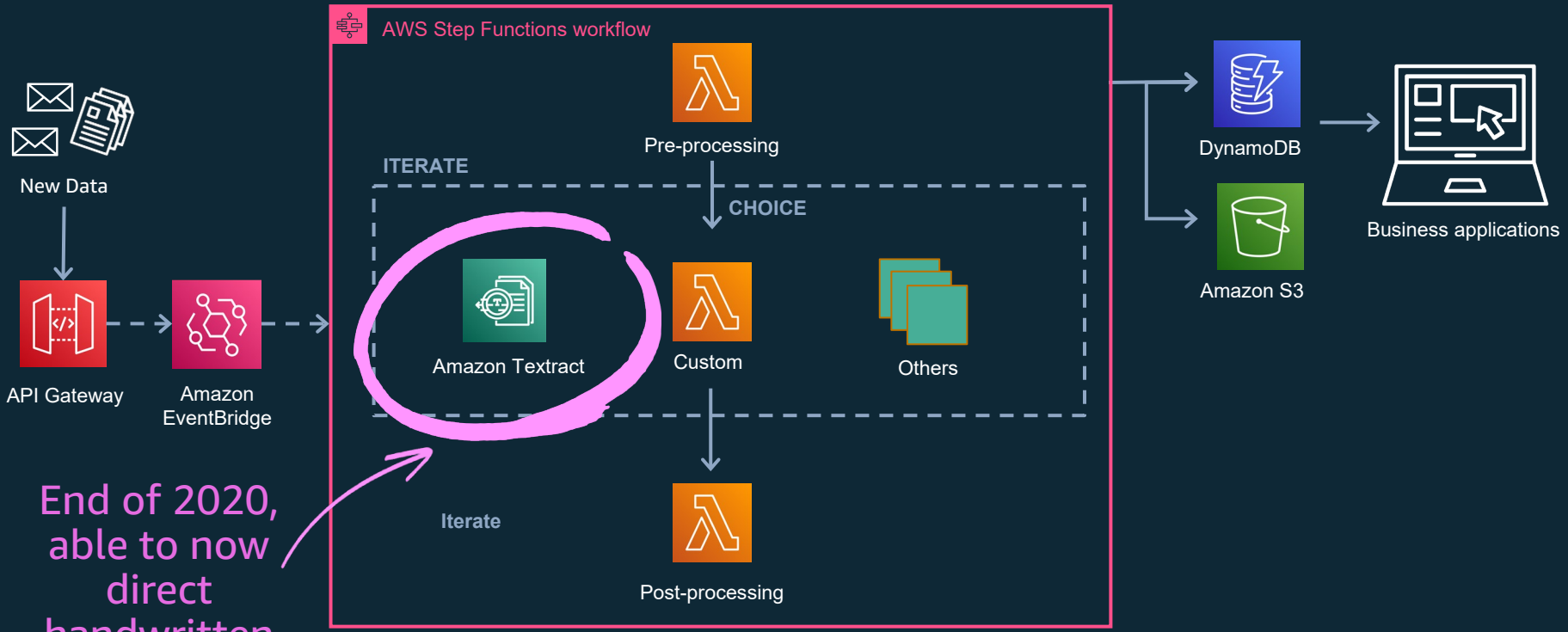
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Real life: Scans and Handwriting!



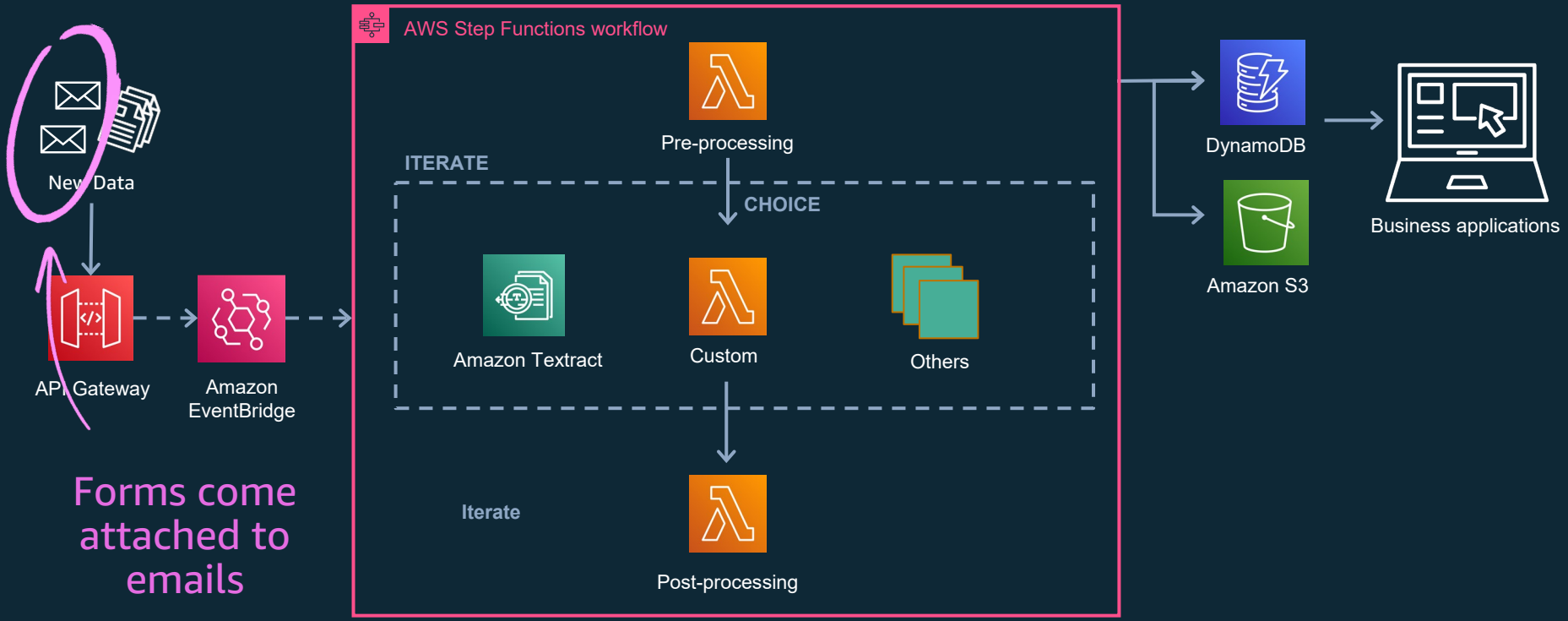
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



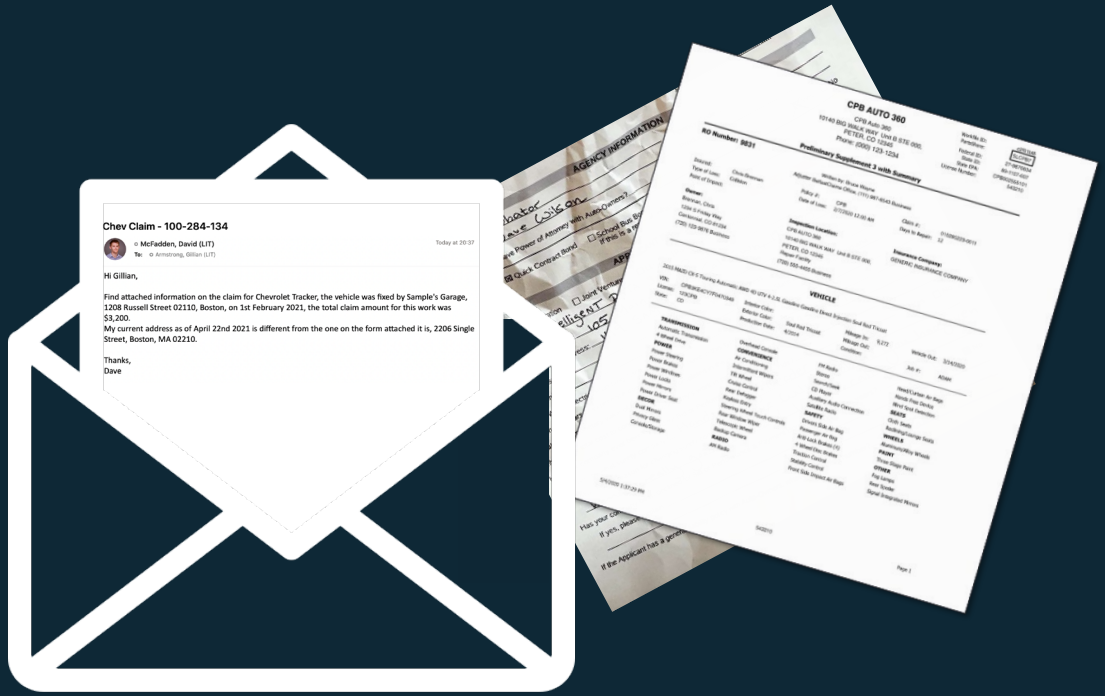
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



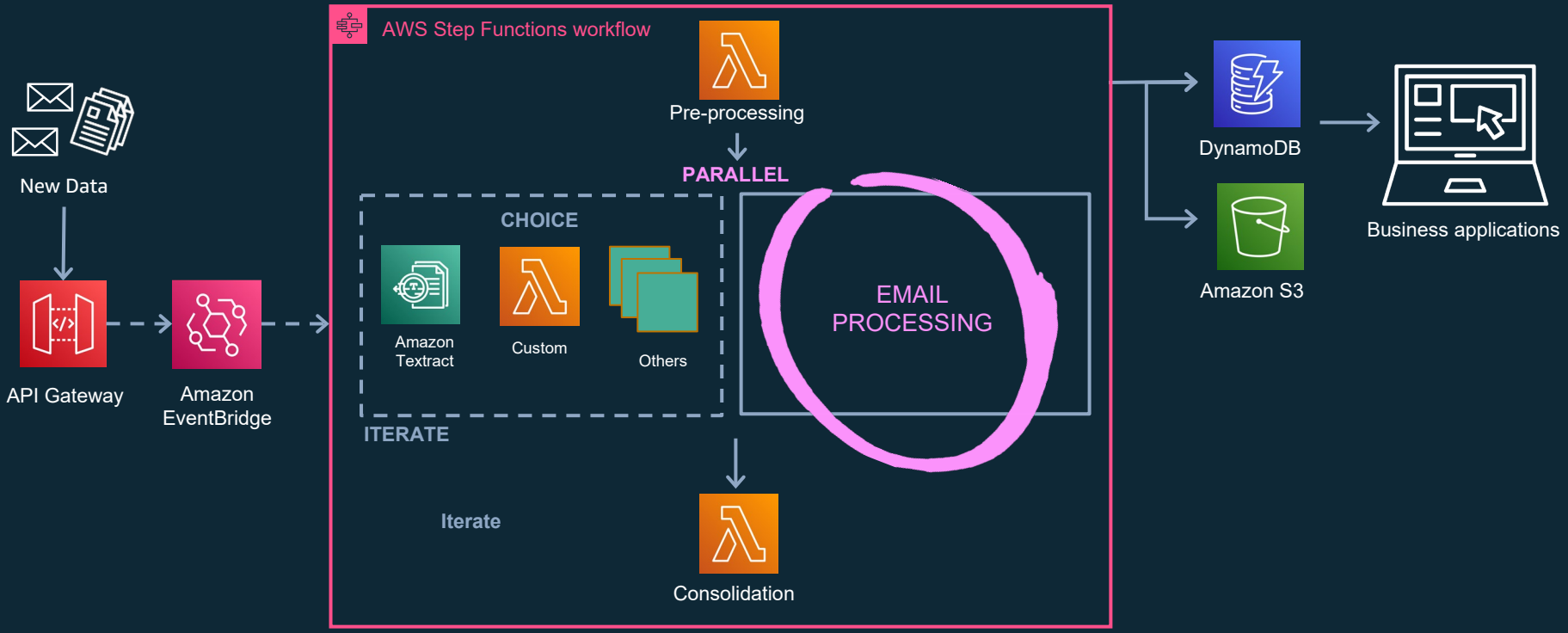
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Emails + Forms!



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

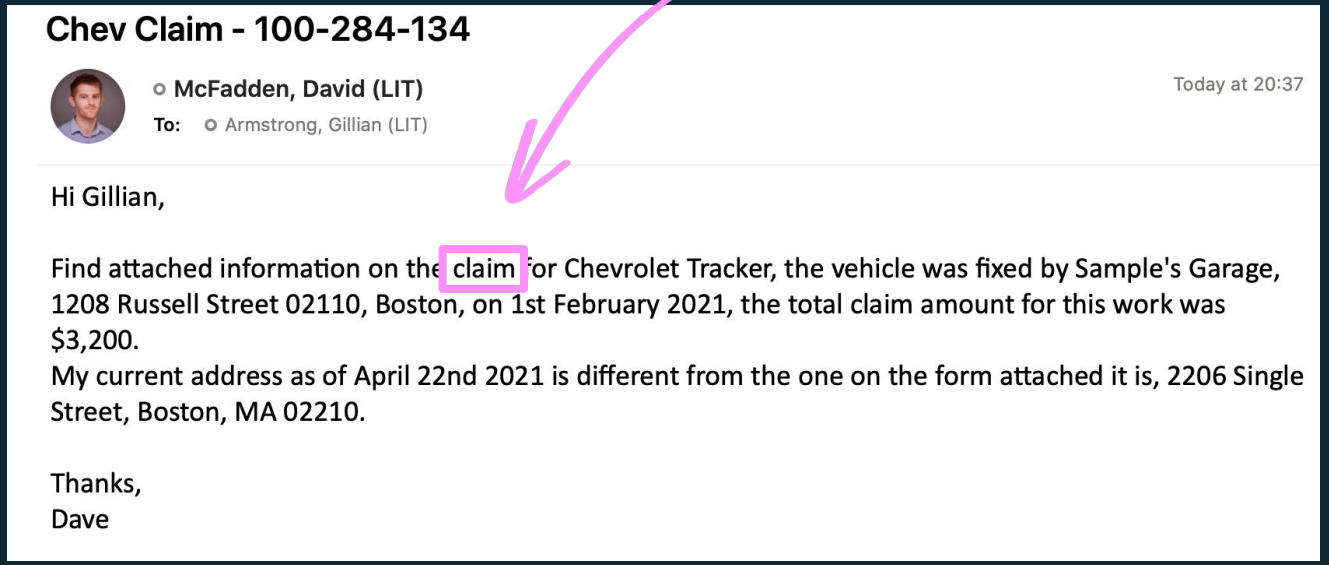
Automated extraction component (simplified)



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Emails

Provides Context



Chev Claim - 100-284-134

McFadden, David (LIT) Today at 20:37

To: Armstrong, Gillian (LIT)

Hi Gillian,

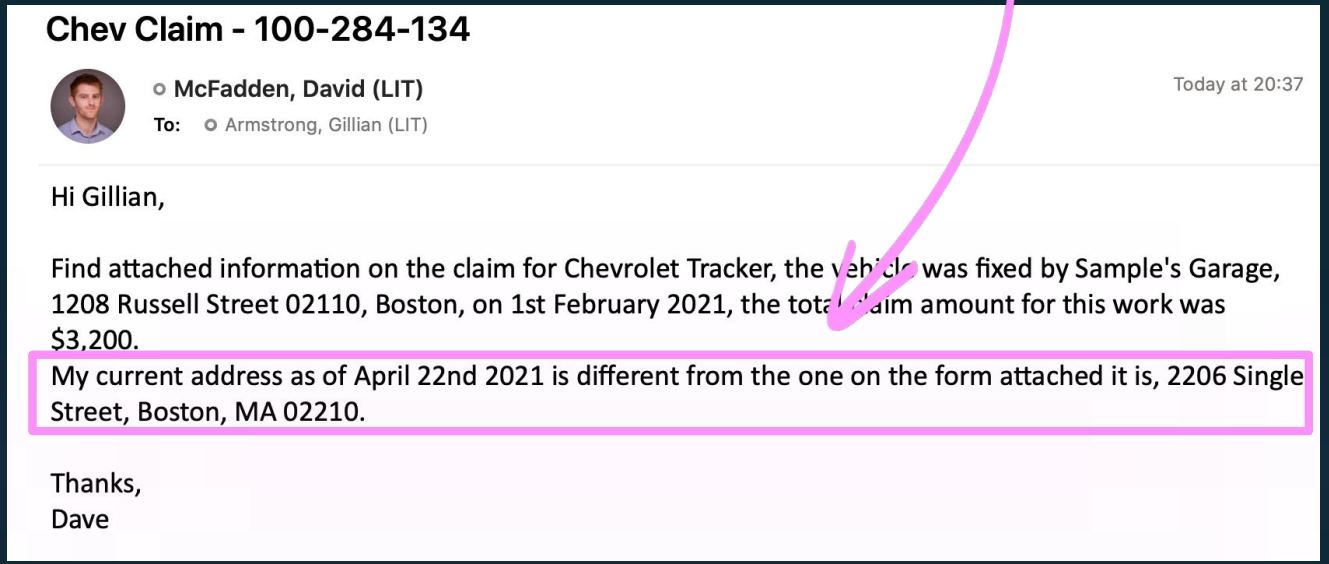
Find attached information on the claim for Chevrolet Tracker, the vehicle was fixed by Sample's Garage, 1208 Russell Street 02110, Boston, on 1st February 2021, the total claim amount for this work was $3,200.
My current address as of April 22nd 2021 is different from the one on the form attached it is, 2206 Single Street, Boston, MA 02210.

Thanks,
Dave

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Emails

Additional Information



Chev Claim - 100-284-134

McFadden, David (LIT)

To: Armstrong, Gillian (LIT)

Today at 20:37

Hi Gillian,

Find attached information on the claim for Chevrolet Tracker, the vehicle was fixed by Sample's Garage, 1208 Russell Street 02110, Boston, on 1st February 2021, the total claim amount for this work was $3,200.

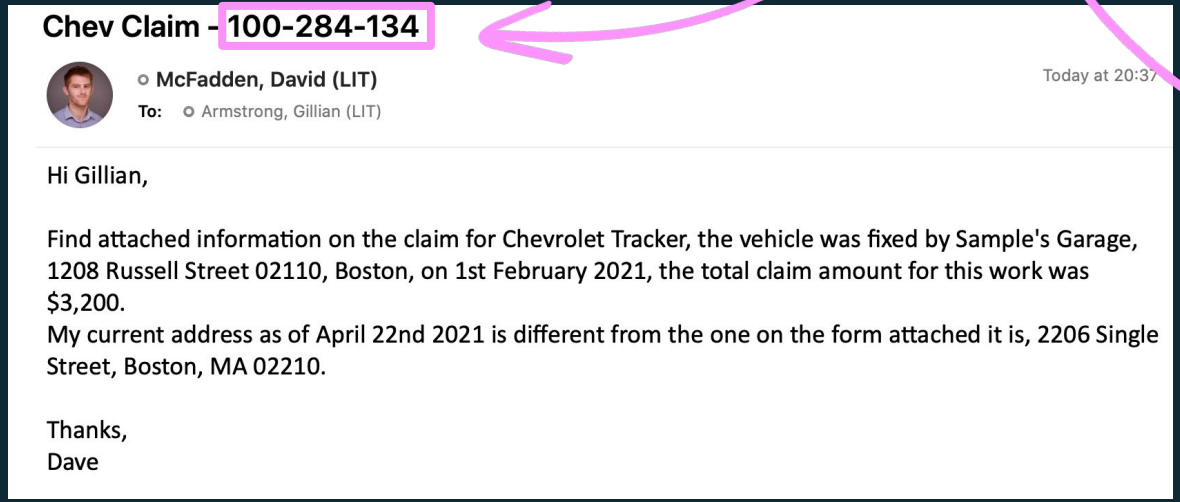
My current address as of April 22nd 2021 is different from the one on the form attached it is, 2206 Single Street, Boston, MA 02210.

Thanks,
Dave

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Emails

Conflicting Information



Chev Claim - 100-284-134

McFadden, David (LIT)

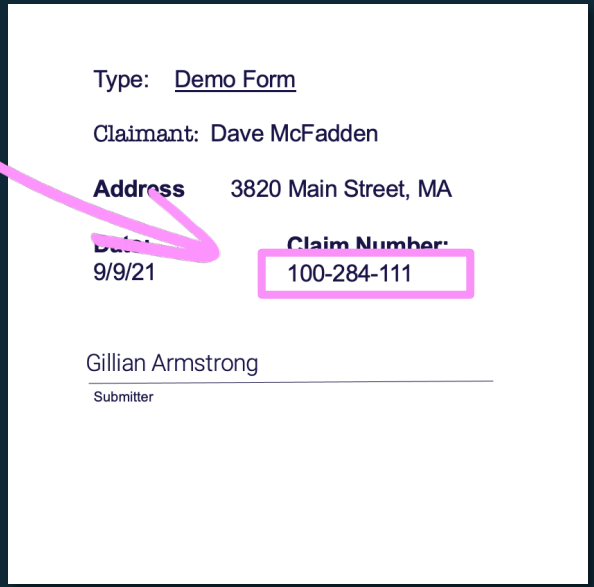
To: Armstrong, Gillian (LIT)

Today at 20:37

Hi Gillian,

Find attached information on the claim for Chevrolet Tracker, the vehicle was fixed by Sample's Garage, 1208 Russell Street 02110, Boston, on 1st February 2021, the total claim amount for this work was $3,200.
My current address as of April 22nd 2021 is different from the one on the form attached it is, 2206 Single Street, Boston, MA 02210.

Thanks,
Dave



Type: Demo Form

Claimant: Dave McFadden

Address 3820 Main Street, MA

9/9/21

Claim Number: 100-284-111

Gillian Armstrong

Submitter

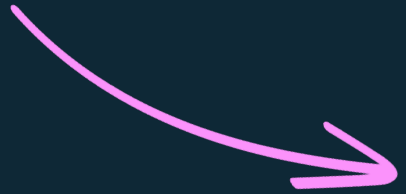
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Comprehend Standard Named Entity Recognition

Hi Gillian,

Find attached information for the claim on 1/1/2021 for Chevrolet Tracker, the vehicle was fixed by Sample's Garage, 1208 Russell Street 02110, Boston, on 1st February 2021, the total claim amount for this work was $3,200.
My current address as of April 22nd 2021 is different from the one on the form attached it is, 2206 Single Street, Boston, MA 02210.

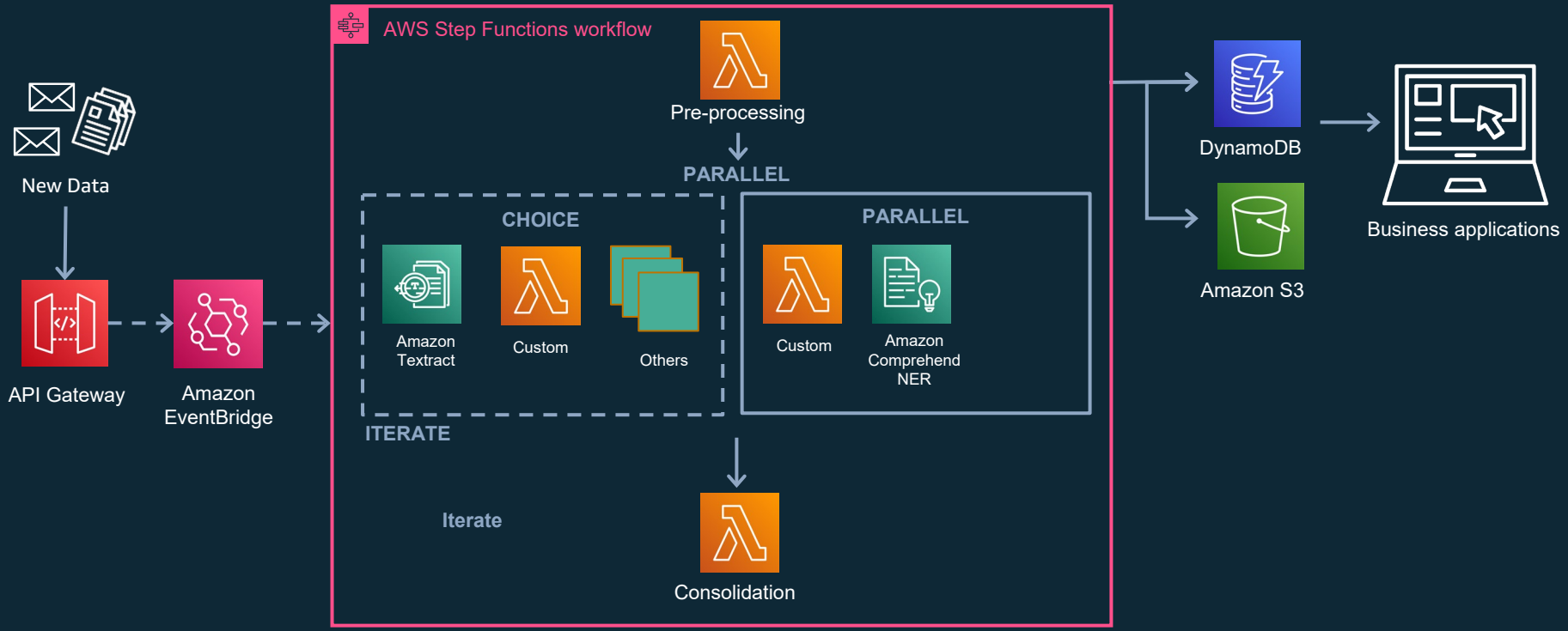
Thanks,
Dave



Entity	Type	Confidence
Gillian	Person	0.99+
1/1/2021	Date	0.99+
Chevrolet	Organization	0.92
Sample's Garage	Organization	0.95
1208 Russell Street 02110, Boston	Location	0.97
1st February 2021	Date	0.99+
$3,200	Quantity	0.99+
April 22nd 2021	Date	0.99+
2206 Single Street, Boston, MA 02210	Location	0.99+
Dave	Person	0.99+

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Comprehend Custom Entity

Hi Gillian,

Find attached information for the claim on 1/1/2021 for Chevrolet Tracker, the vehicle was fixed by Sample's Garage, 1208 Russell Street 02110, Boston, on 1st February 2021, the total claim amount for this work was $3,200.
My current address as of April 22nd 2021 is different from the one on the form attached it is, 2206 Single Street, Boston, MA 02210.

Thanks,
Dave

Claim Date →

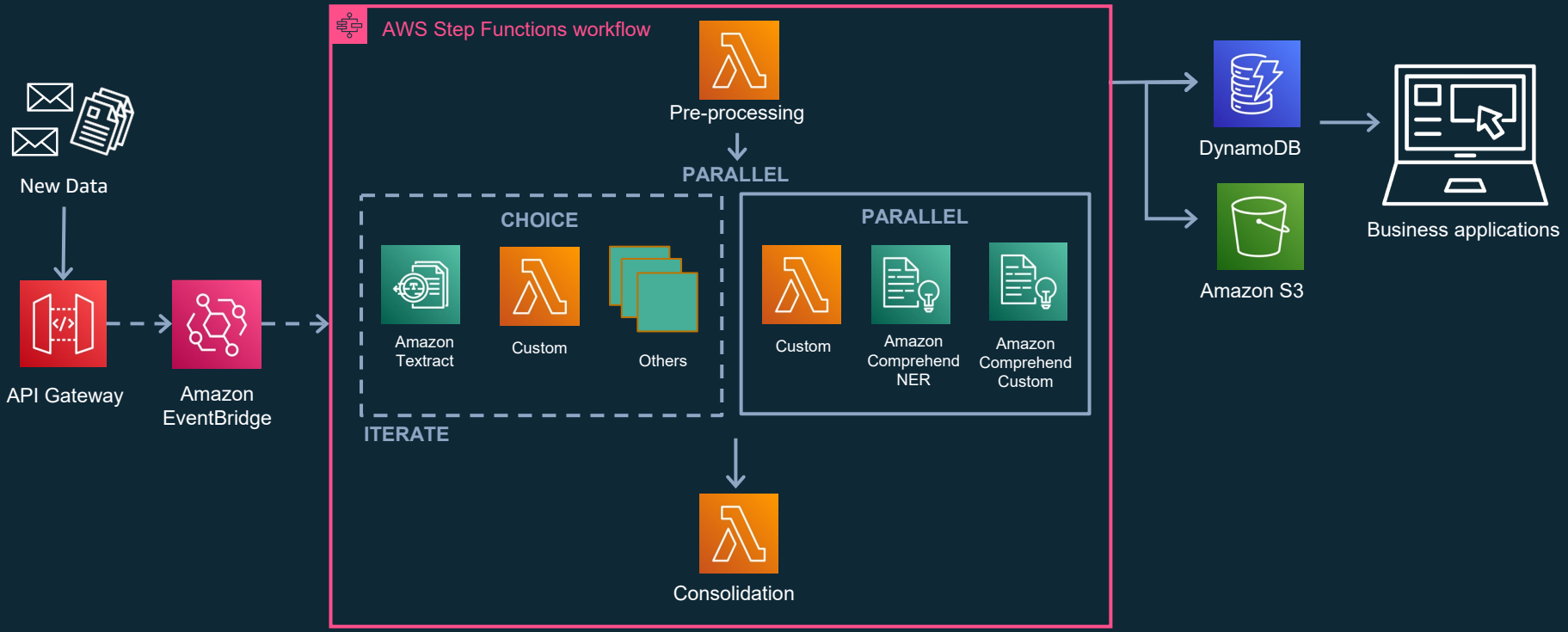
Repair Date →

Address Change →

Entity	Type	Confidence
Gillian	Person	0.99+
1/1/2021	Date	0.99+
Chevrolet	Organization	0.92
Sample's Garage	Organization	0.95
1208 Russell Street 02110, Boston	Location	0.97
1st February 2021	Date	0.99+
$3,200	Quantity	0.99+
April 22nd 2021	Date	0.99+
2206 Single Street, Boston, MA 02210	Location	0.99+
Dave	Person	0.99+

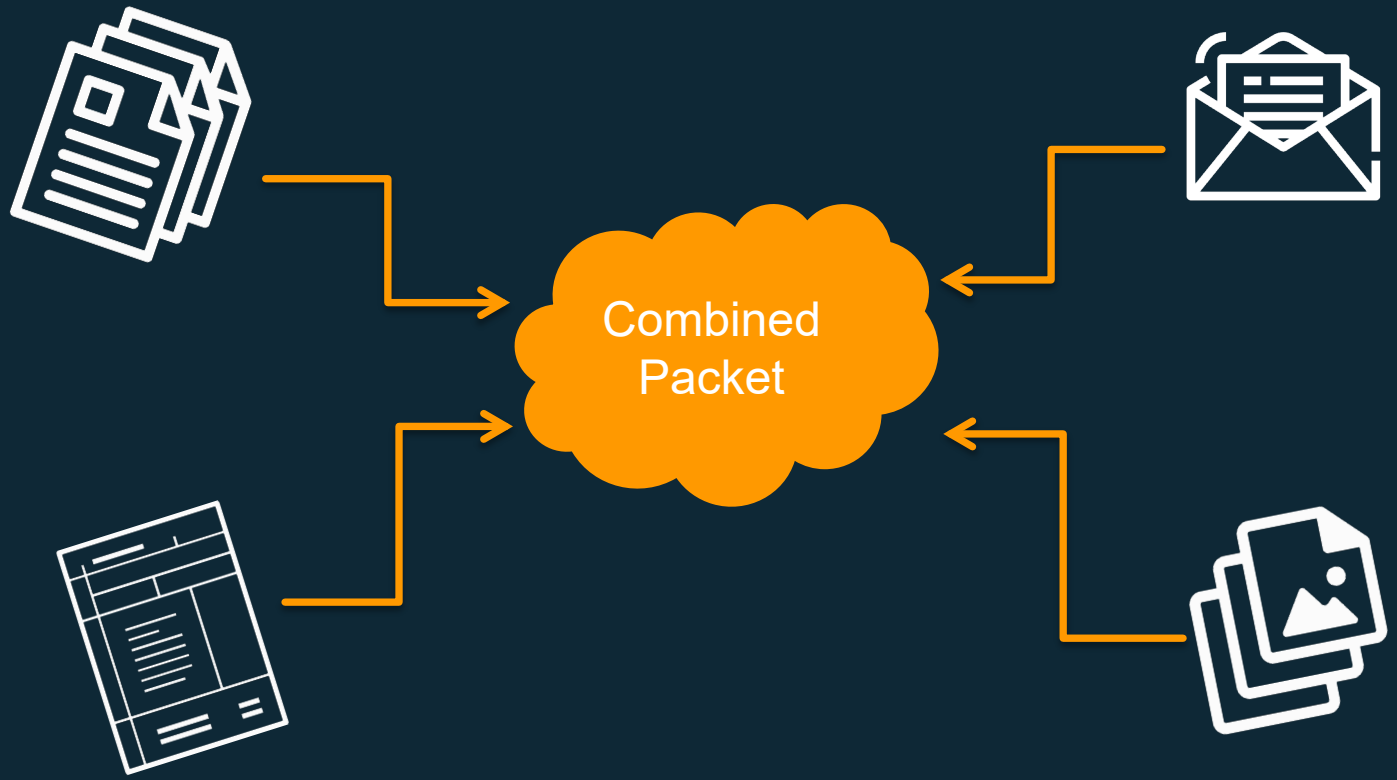
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Automated extraction component (simplified)



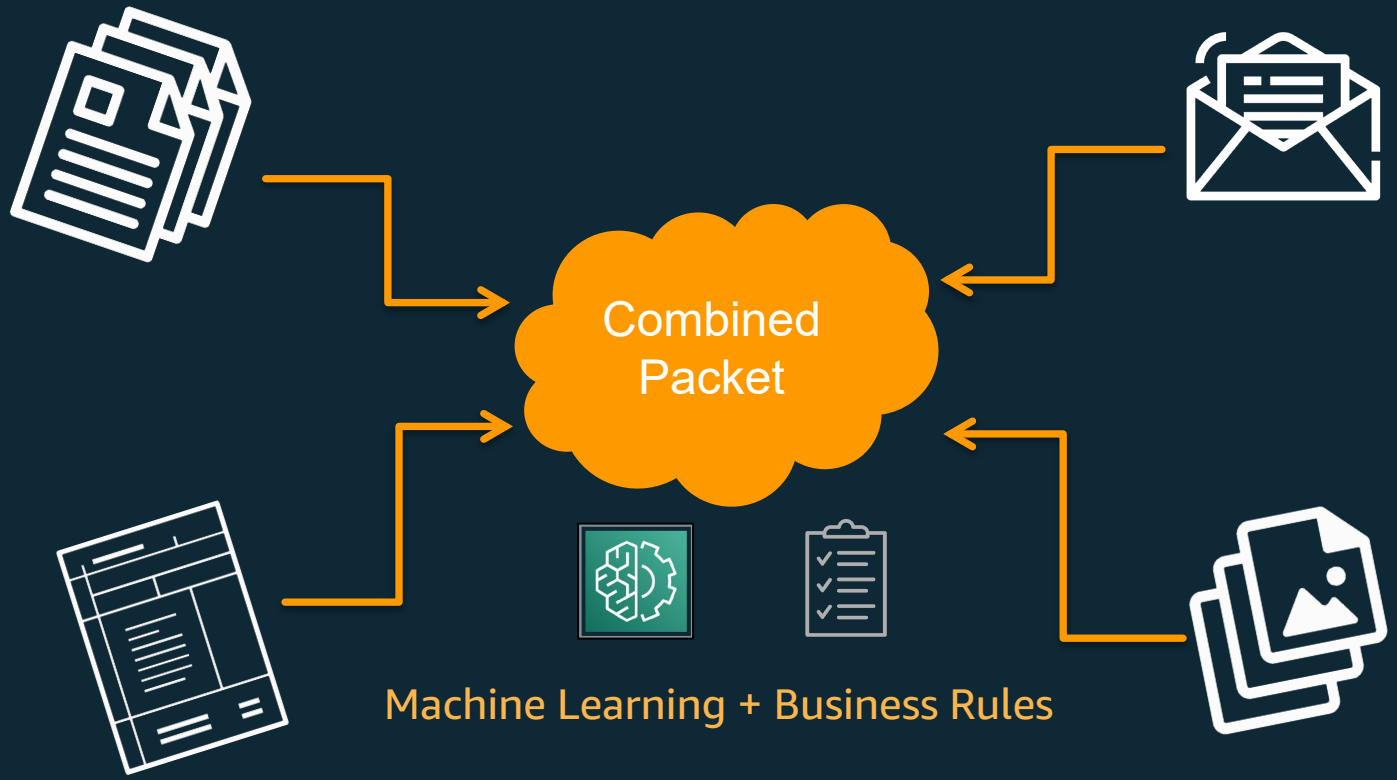
© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Post Processing / Consolidation



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Post Processing / Consolidation



© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Intelligent Data Processing

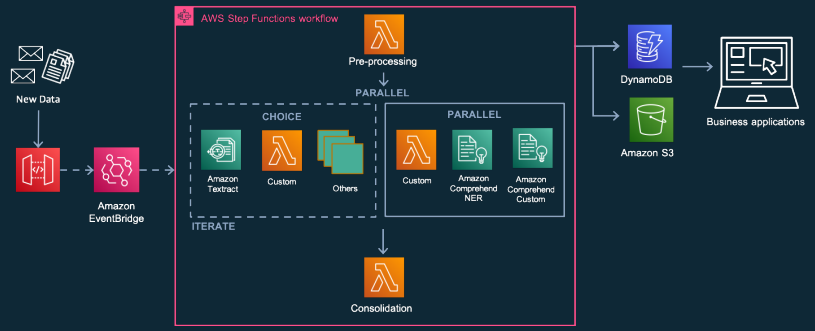
Combining Unstructured Data…

…to create Structured Data that can be used for automation, analysis, and connected to other existing data

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Where are we now?

- Extensible and flexible Serverless Platform in place, orchestrated with AWS Services, and with AWS ML Services providing key extraction options.



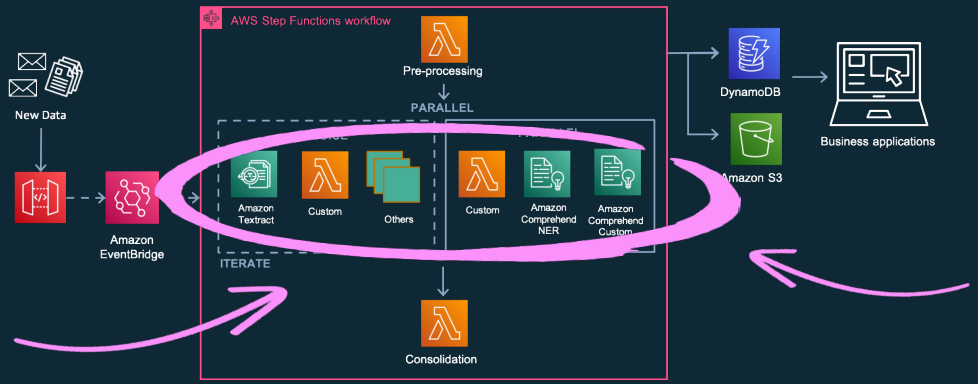
- 40% efficiency gains in data extraction and input.
- New use-cases being onboarded into the system from across the company.

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Looking to the Future…

- Use learnings to continuously evolve and improve extraction methods and overall platform
- Expand to new types of data and even more ambitious use-cases

Continue to improve existing extraction methods



Continue to expand with new custom models and integration of additional AWS ML Services

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

How we can help

Run a POC

- Run a quick POC with ready-to-deploy Document Understanding Solution
- https://aws.amazon.com/solutions/implementations/document-understanding-solution/

AWS Professional Services Team

- Identify key areas where machine learning could help you with an AWS AI expert
- Let AWS experts build POCs and custom solutions for your use case

Solution architects

- Engage with a solution architect dedicated to your account
- Brainstorm use cases and get support for solution implementation
- https://aws.amazon.com/contact-us/

AWS Sales & Account teams

- Contact sales if you have any queries about building intelligent document processing workflows with AWS

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.