



**19 August 2021**

# Building a Document Processing Solution with Amazon Textract, Database, Data Lakes and Analytics on AWS

Ryan Hendriks

Principal Solution Architect – Data Modernization  
Amazon Web Services



# Agenda

- Document Processing Challenges
- Introduction to Amazon Textract  
**Demo: Drag and Drop**
- Emerging use cases
- Textract Solution Architecture  
**Demo: Unstructured to Structured | Textract API → Lake House → QuickSight**
- Takeaways & Next Steps

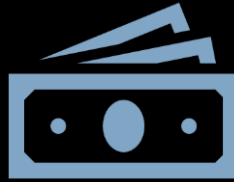
# Document Processing Challenges



Time



Error



Cost



Scale



Customer  
Experience

# Document Processing Challenges

## Digital Forms



## Onboarding documents

Identification cards, registration forms, proof documents

## Handwritten



## Financial and transactional documents

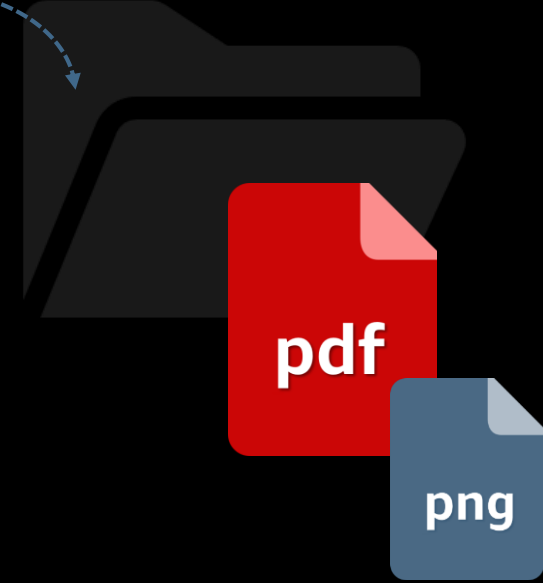
Purchase orders, contracts, pay slips, invoices/receipts, bank statements

## Typed or Printed



## Complex documents and archives

Reports, policies, product catalogs, engineering diagrams, handwritten documents



Difficult to **query**

Difficult to **analyze**

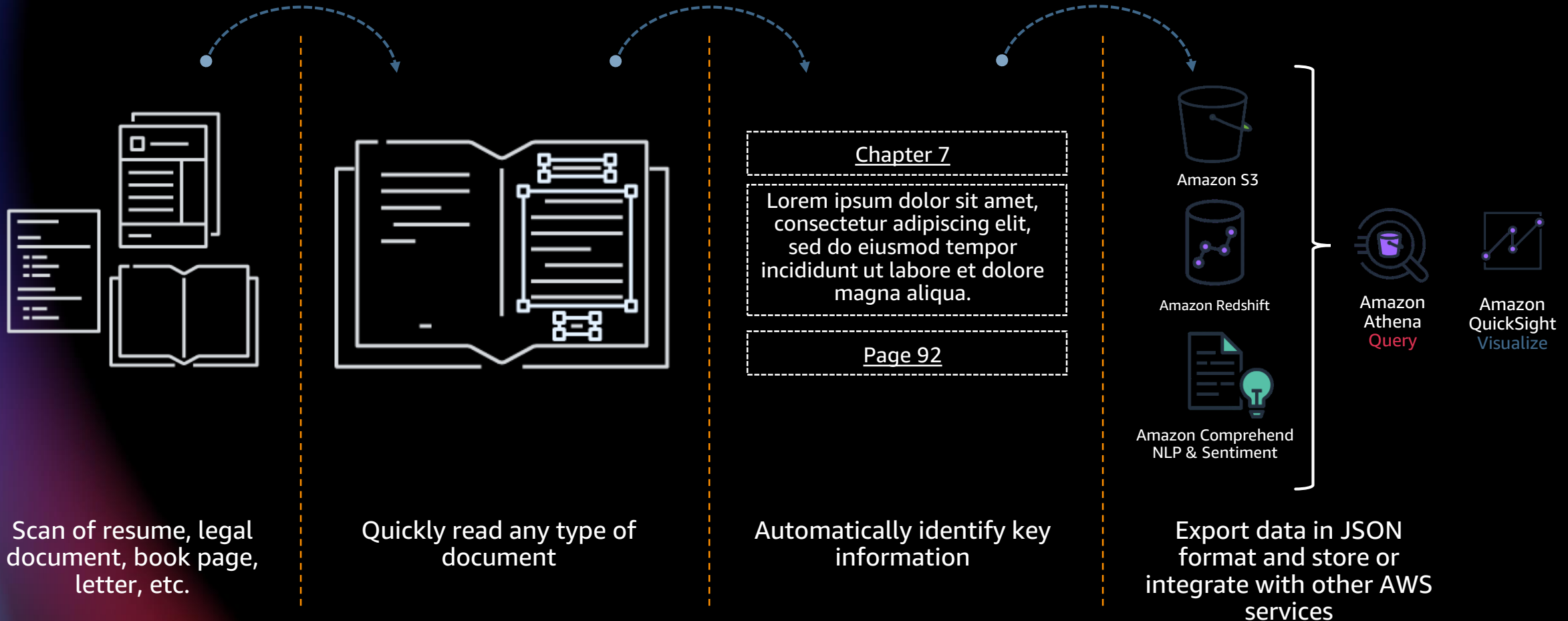
Difficult to **visualize**

**at Scale**



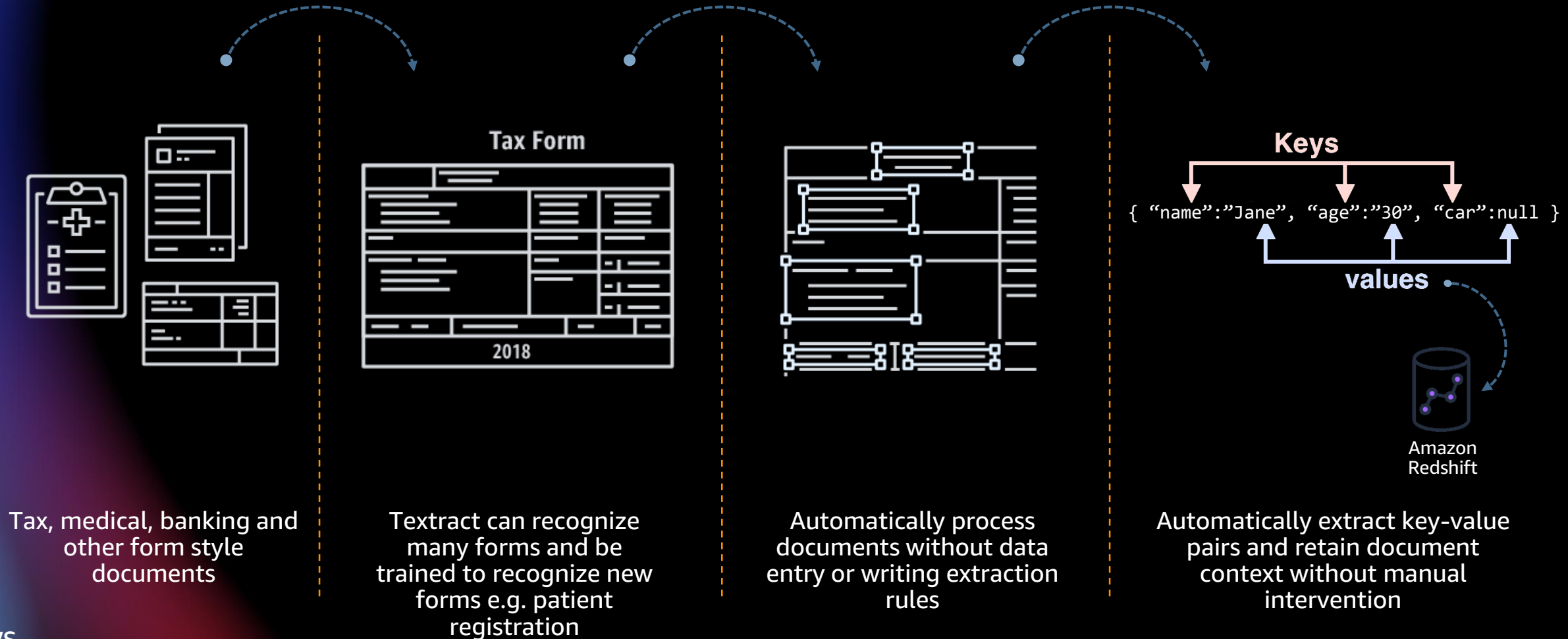
# Introduction to Amazon Textract

## OPTICAL CHARACTER RECOGNITION



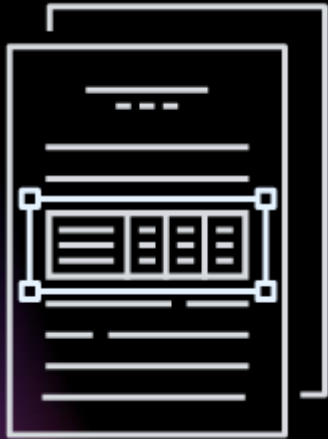
# Introduction to Amazon Textract

## FORM EXTRACTION

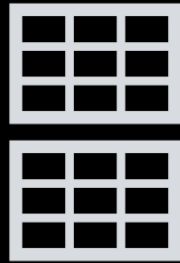


# Introduction to Amazon Textract

## TABLE EXTRACTION



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



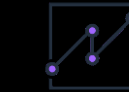
Textract preserves the tabular structure of extracted data so text remains grouped in each cell



With the tabular format of the data intact, easily upload data to database



With structured data in a database easily build compelling visualizations from the data





# Introduction to Amazon Textract

DRAG AND DROP | WEB USER INTERFACE

## Demo

# Emerging Use Cases



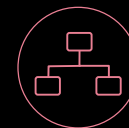
## Intelligent Search

Digital searchable catalogues  
Enhance data lake with document data



## Compliance and Control

Identify and redacts personally identifiable information (PII) or other sensitive data.



## Process Automation

Enable automatic data ingestion in CRMs, ERPs

### Solutions

### Industry



## Financial Services

Financial **forms** like mortgage applications can contain critical business information like **rates**, applicant **names** and important **tax information** which needs to be extracted and analyzed.

With Amazon Textract, you can **extract information from scanned documents, tables and forms**, and process applications in minutes to provide your customers a quick response.



## Healthcare and Lifesciences

Amazon Textract can scan thousands of **healthcare and insurance forms**, extract the information from within those forms and keep the information **organized** and in its original context, saving you from **manually reviewing the output**.

Better serve your patients and insurers by extracting **important patient data** from **health intake forms, insurance claims, and pre-authorization forms**.



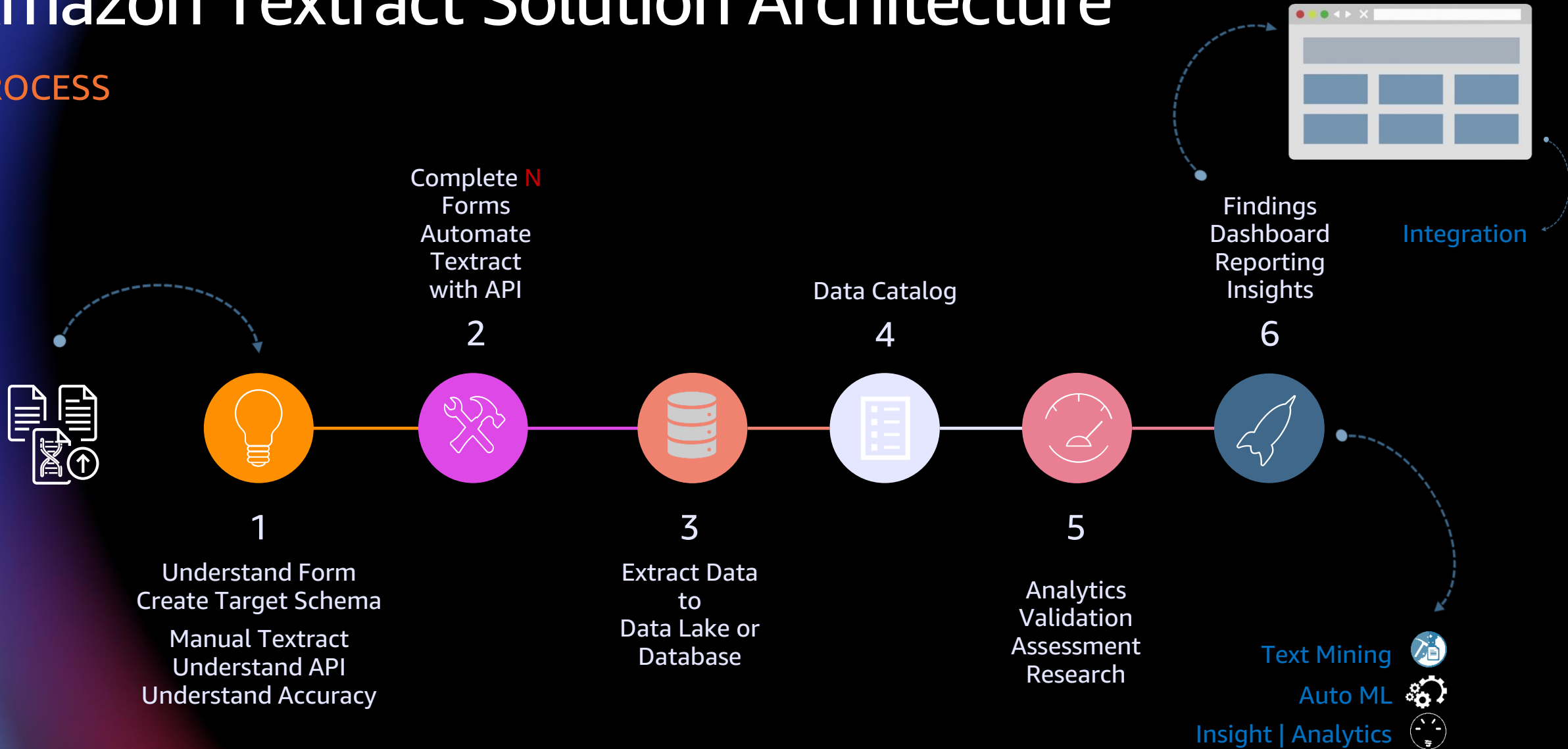
## Public Sector

Processing government related forms like small business loans, **federal tax forms** or **business applications** takes **thousands of manual hours** to extract the relevant and important data.

Amazon Textract can extract all the data from **images, PDFs or scanned documents** using Optical Character Recognition (OCR). Textract not only identifies each character, word and letter, but also the contents of form fields and information stored in tables with high accuracy.

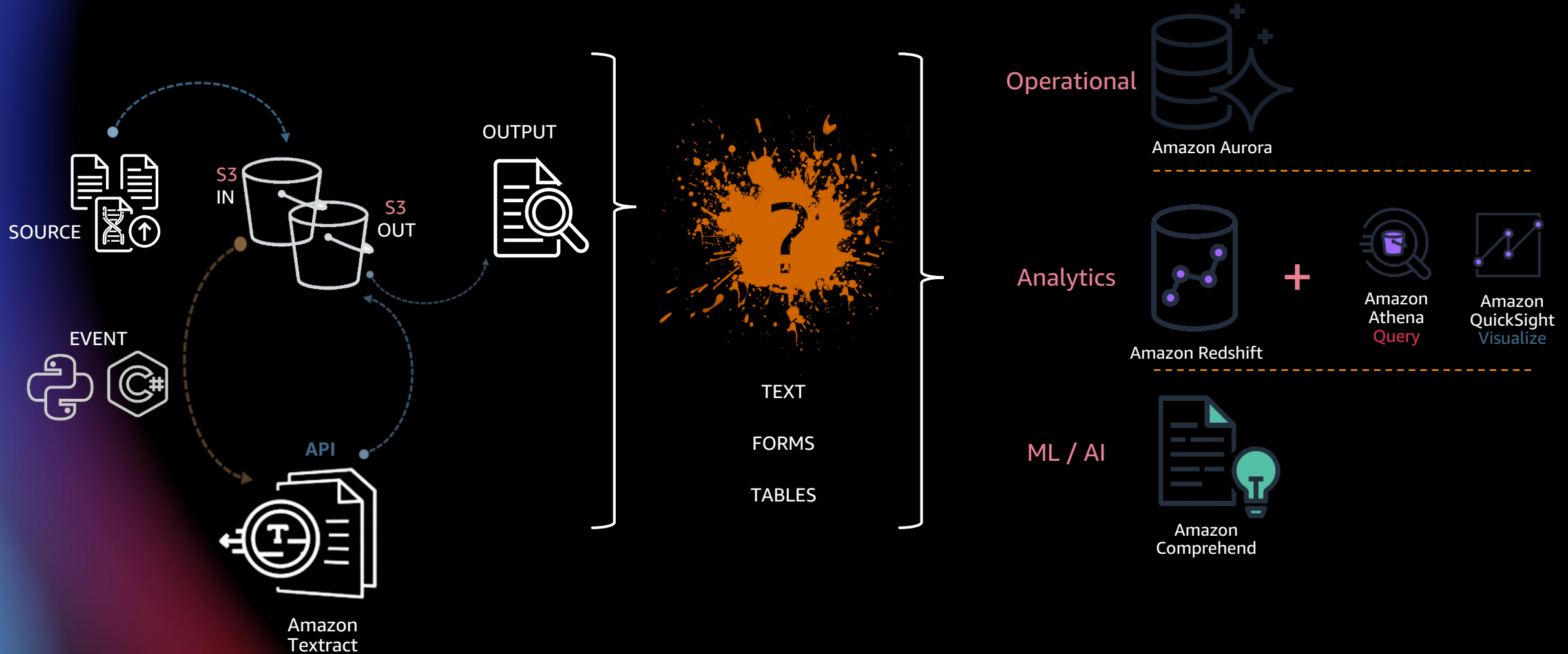
# Amazon Textract Solution Architecture

## PROCESS



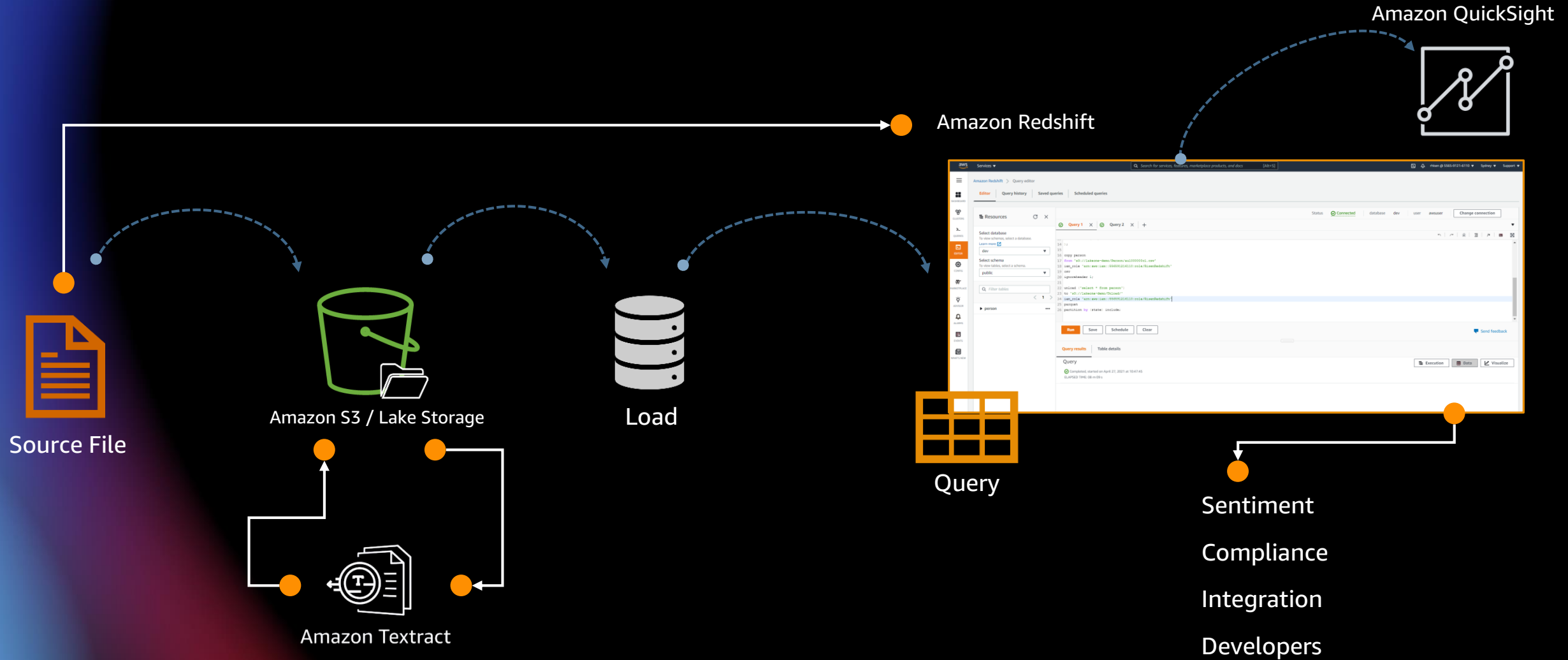
# Amazon Textract Solution Architecture Example

## SERVICES



# Introduction to Amazon Textract Demo

DEMO: UNSTRUCTURED TO STRUCTURED | TEXTTRACT API → LAKE HOUSE → QUICKSIGHT

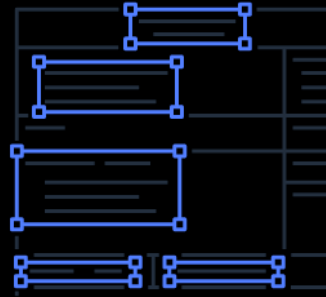


# Takeaways



Do more  
with less

Operational efficiency



Complex documents  
can be automated

Leverage AI

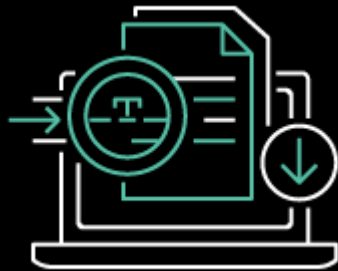


Consistency and  
Customer experience

Unstructured to Structured  
and  
Accessible Analytics

# Next Steps

## Try Amazon Textract for Free



<https://aws.amazon.com/textract>

## Download code to automate



<https://github.com/aws-samples/amazon-textract-code-samples>

# Additional resources

- [Automate data extraction and analytics from documents with machine learning](#)  
- 2 min video
- [DUS – Document Understanding Solution](#)
- [Getting started with Amazon Textract / Amazon Comprehend](#)
- [Amazon Textract / Amazon Comprehend pricing](#)
- [Building an end-to-end intelligent document processing solution using AWS](#)  
- Blog post



# Visit the AWS Data Resource Hub

Dive deeper with these resources, get inspired and learn how you can use data to make better decisions and innovate faster.

- Building a winning data strategy
- The new leadership mindset for data & analytics
- Harness data to reinvent your organization
- Put your data to work with a modern analytics approach
- Breaking free from on-premises database constraints
- Cloud storage adoption: From cost optimization to agility & innovation
- A strategic playbook for data, analytics, and machine learning
- ... and more!



<https://tinyurl.com/aws-data-resource>

Visit resource hub



# AWS Training and Certification

## Empower your teams with comprehensive training

By building skills with AWS Training and Certification, businesses and individuals can see the bigger picture understanding the reasoning behind every data point. As training progresses and teams become data-fluent, previously hidden insights come into view.

Build data skills to  
**unlock any insight**

### Leverage free digital training

Learn how to harness the world's most valuable resource: data. Access digital and virtual instructor-led courses on data analytics and databases built by the experts at AWS and start your learning journey to become data-driven.

[Take a digital course »](#)



### Get certified

Earn industry-recognized credibility and set tangible goals for success with industry-recognized certifications, like *AWS Certified Data Analytics – Specialty*.

[Learn more »](#)



### Ramp-up your skills

Deep dive into new topics and focus on knowledge gaps at your own pace with the *AWS Ramp-Up Guide: Database* and *AWS Ramp-Up Guide: Data Analytics*. With a wide range of whitepapers, blog posts, videos, webinars and peer resources available for data professionals to leverage for independent learning.

[Download ramp-up guides »](#)

# Thank you for attending AWS Innovate – Data Edition

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-apj-marketing@amazon.com](mailto:aws-apj-marketing@amazon.com)



[twitter.com/AWSCloud](https://twitter.com/AWSCloud)



[facebook.com/AmazonWebServices](https://facebook.com/AmazonWebServices)



[youtube.com/user/AmazonWebServices](https://youtube.com/user/AmazonWebServices)



[slideshare.net/AmazonWebServices](https://slideshare.net/AmazonWebServices)



[twitch.tv/aws](https://twitch.tv/aws)

# Thank you!

Ryan Hendriks  
Principal Solution Architect – Data Modernization  
Amazon Web Services

