



# aws INNOVATE

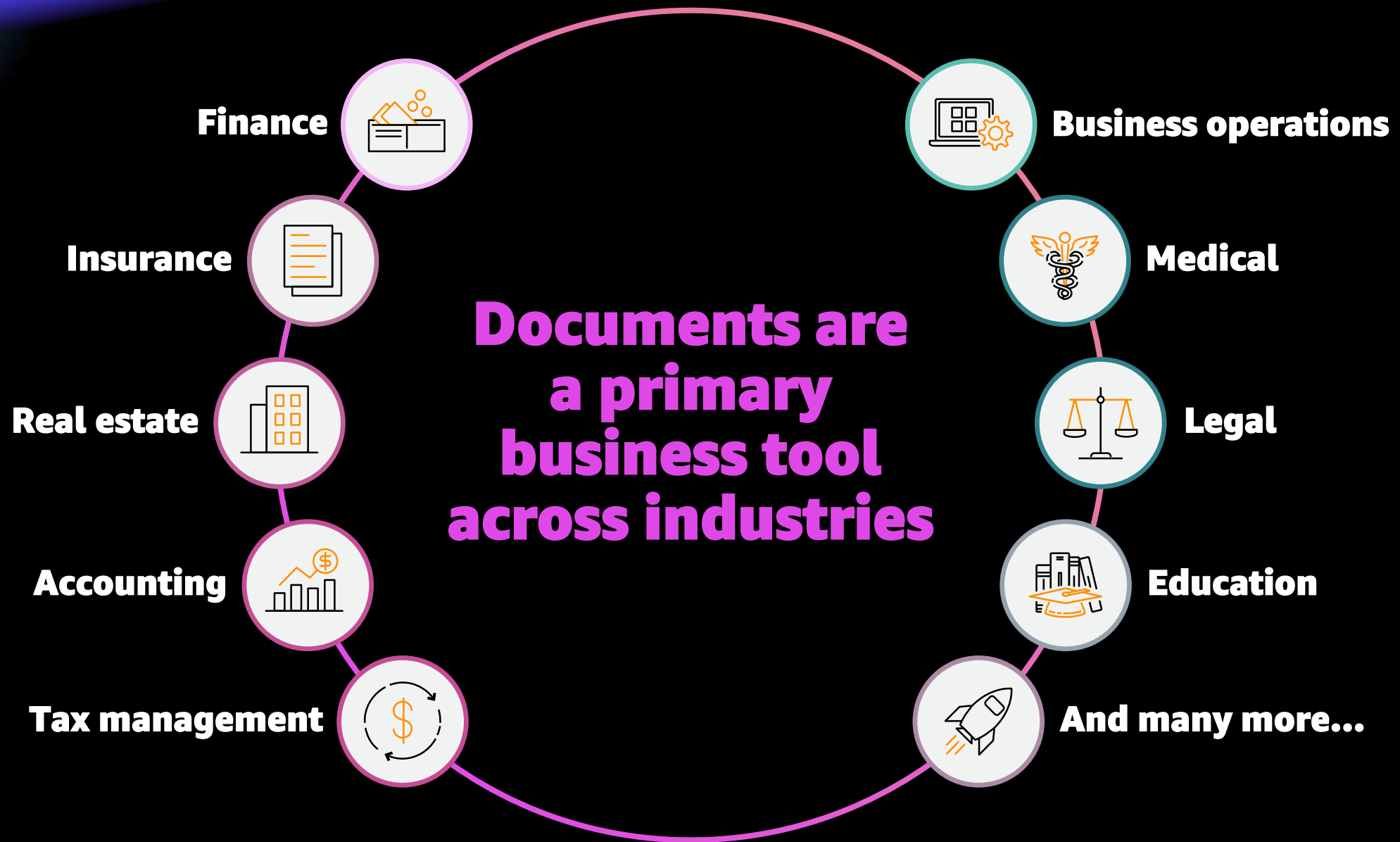
AI/ML EDITION

24 February 2022

# Intelligent unstructured document processing

Santhosh Urukonda  
Prototyping Architect, AISPL





# How documents are processed today



## Manual processing

- ⊗ Expensive
- ⊗ Error prone
- ⊗ Time consuming



## Traditional Optical Character Recognition (OCR)

- ⊗ Simple documents only
- ⊗ Error prone
- ⊗ Dump of text

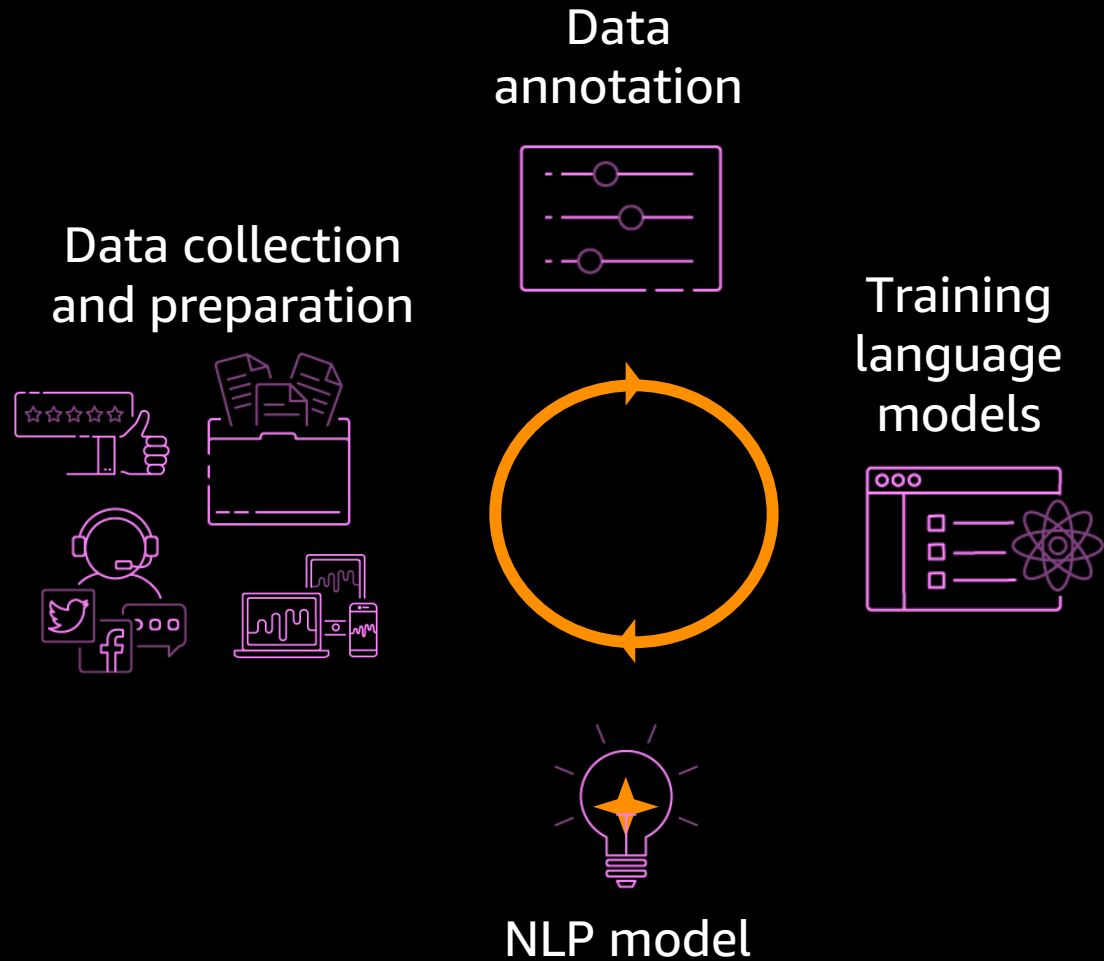


## Rules and template-based extraction

- ⊗ Limited by OCR accuracy
- ⊗ Development and management overhead
- ⊗ Templates are brittle

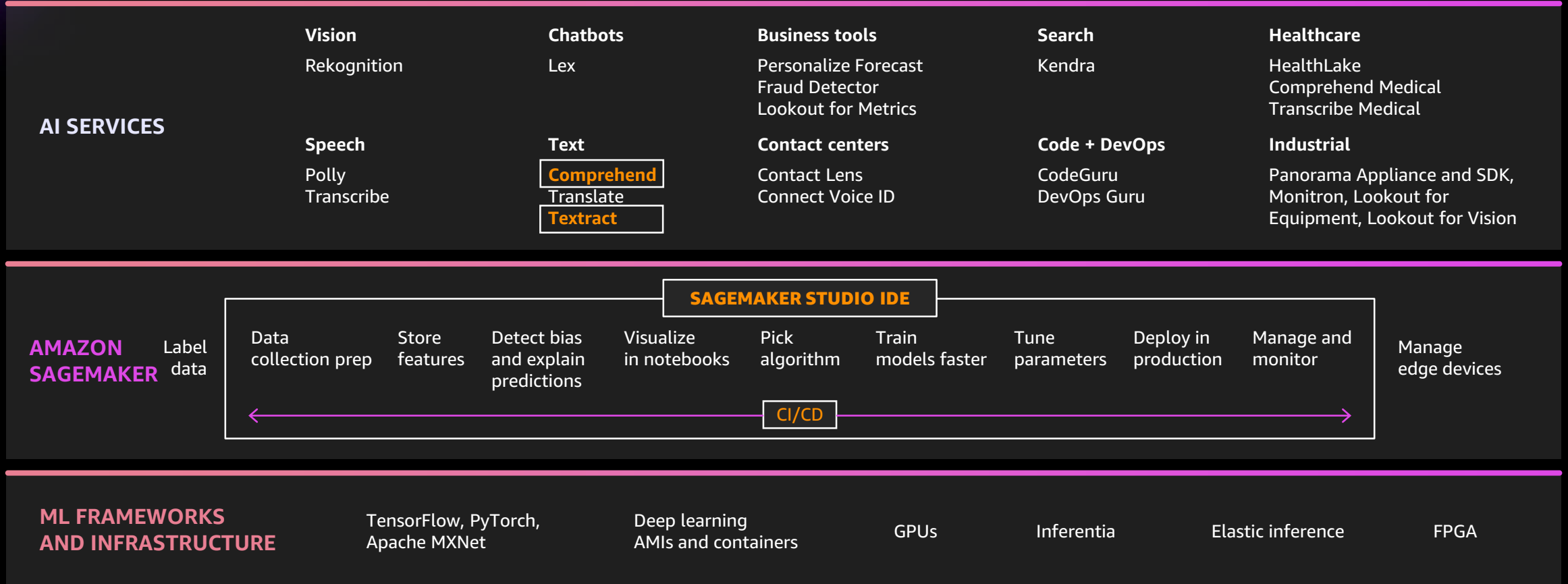
# Customer challenges with building NLP models

- Machine learning skills
- Labeled training data
- Model training effort
- Model management in production

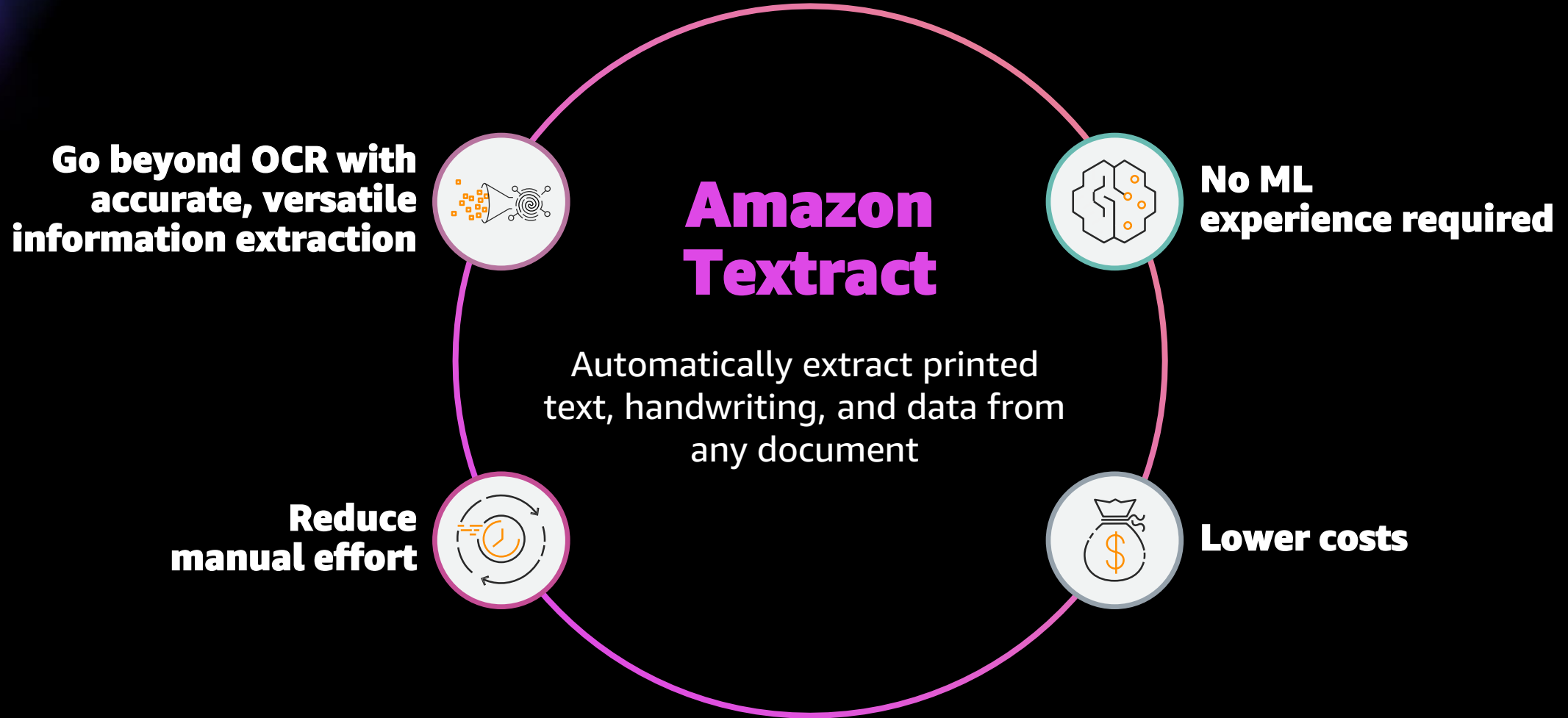


# The AWS ML stack

BROADEST AND MOST COMPLETE SET OF MACHINE LEARNING CAPABILITIES



# Amazon Textract



# Amazon Comprehend



Documents

Email, chat,  
social, phone  
calls and more



Amazon  
Comprehend

Automatically  
extract insights  
from text



Entities

+ Custom Entities



Key Phrases



PII

(Personally Identifiable  
Information)



Sentiment



Document  
Classification



Topics



Language



Syntax

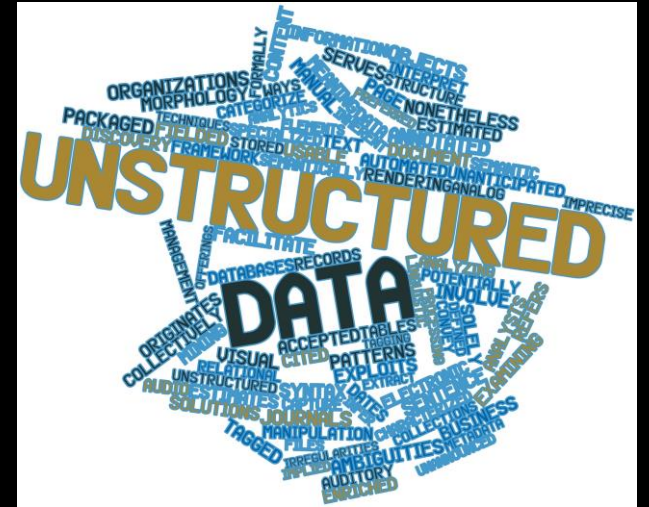


Events



# Use case

- Process unstructured documents like product reviews
- Data is not in fixed/organized format
- Multiple file types (doc, pdf, image etc.)
- Technical jargon specific to product
- Product feature comparison
- Overall sentiment of the product



# AWS services used

1



Extract text from documents

Amazon Textract

2



Entity detection & sentiment

Amazon Comprehend

3



Invoke other AWS services &  
run custom rules

AWS Lambda

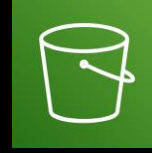
4



Orchestrate workflow

AWS Step Functions

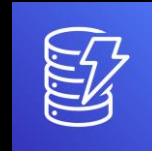
5



Document storage

Amazon S3

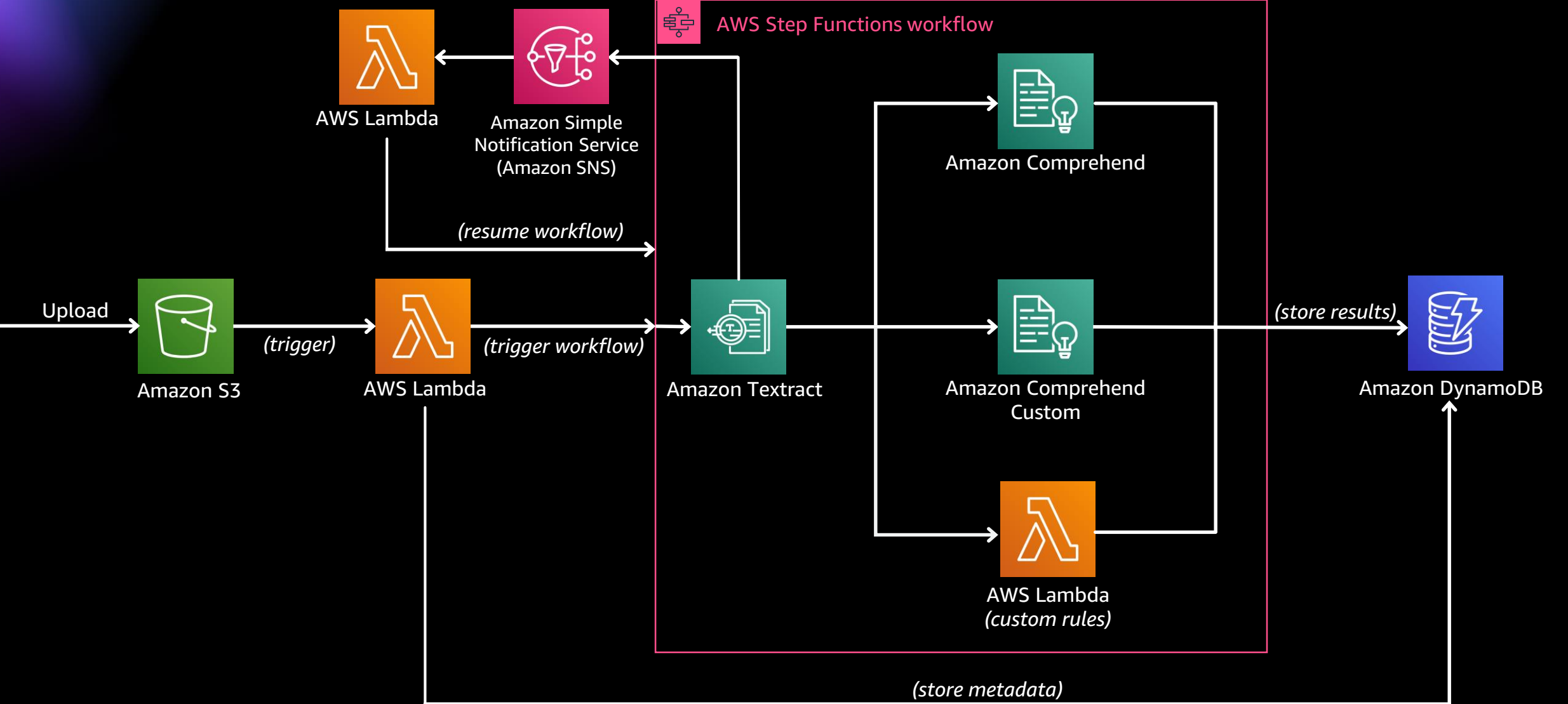
6



Store results & job tracking

Amazon DynamoDB

# Architecture



# Demo

# Recap

- Unstructured document processing challenges
- AWS AI services (Amazon Textract & Amazon Comprehend)
- Use case
- System architecture
- Solution walkthrough

# Additional resources

## Amazon Textract Blogs



<https://go.aws/3259Sm8>

## Amazon Comprehend Blogs



<https://go.aws/3qxDVwm>

# Visit the AI & Machine Learning resource hub for more resources

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

- The machine learning journey e-book
- 7 leading machine learning use cases e-book
- A strategic playbook for data, analytics, and machine learning e-book
- Accelerate machine learning innovation with the right cloud services & infrastructure e-book
- Choosing the right compute infrastructure for machine learning e-book
- Improving service and reducing costs in contact centers e-book
- Why ML is essential in your fight against online fraud e-book
- ... and more!



<https://bit.ly/3mwi59V>

Visit resource hub

# AWS Machine Learning (ML) Training and Certification



## 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.

[aws.training/machinelearning](https://aws.training/machinelearning)



## Flexibility to learn your way

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

[explore.skillbuilder.aws/learn](https://explore.skillbuilder.aws/learn)



## Validate your expertise

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

[aws.amazon.com/certification](https://aws.amazon.com/certification)



# Thank you for attending AWS Innovate – AI/ML 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!

Santhosh Urukonda

