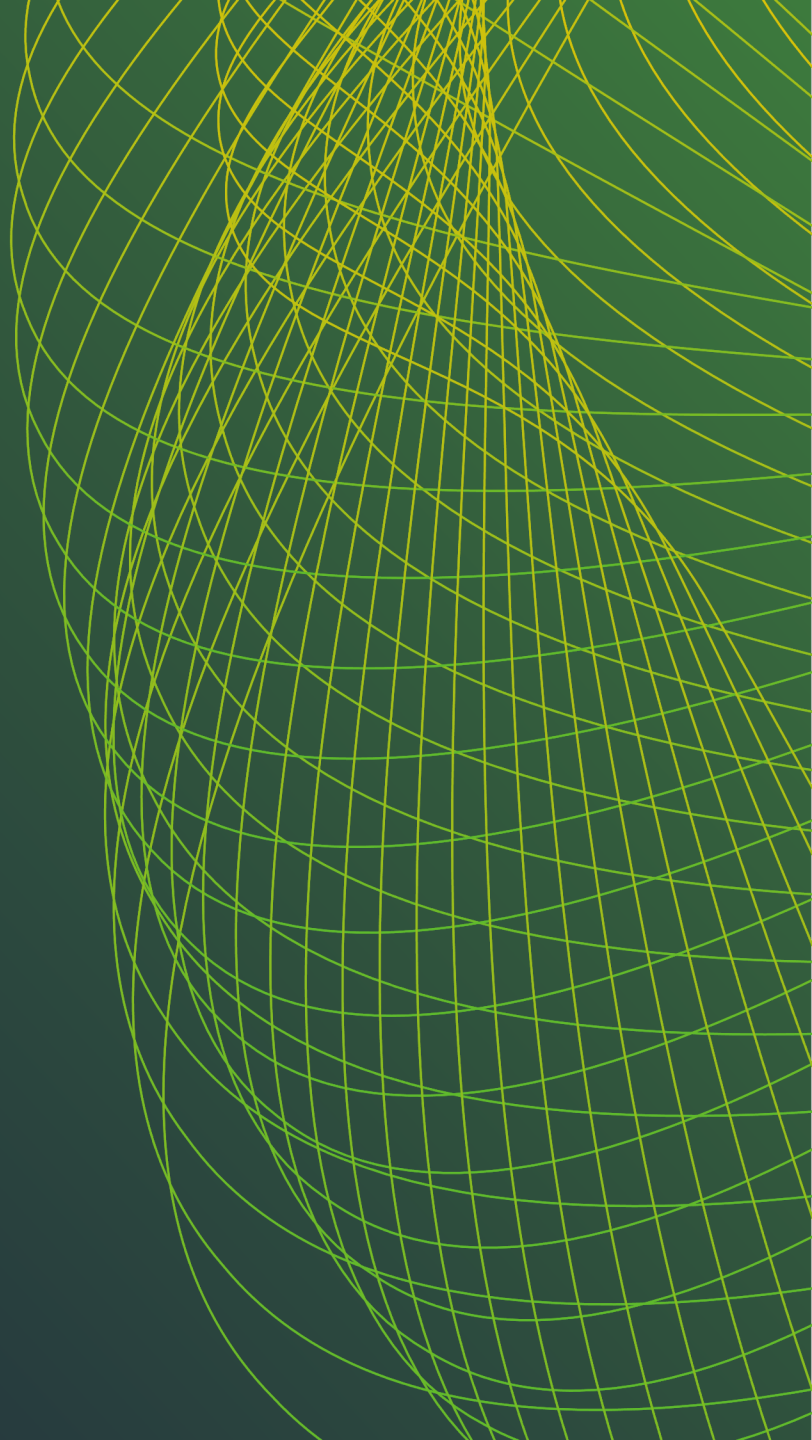


FINANCIAL SERVICES
CLOUD SYMPOSIUM

MAY 4, 2021

FINANCIAL SERVICES
CLOUD SYMPOSIUM

FSI101

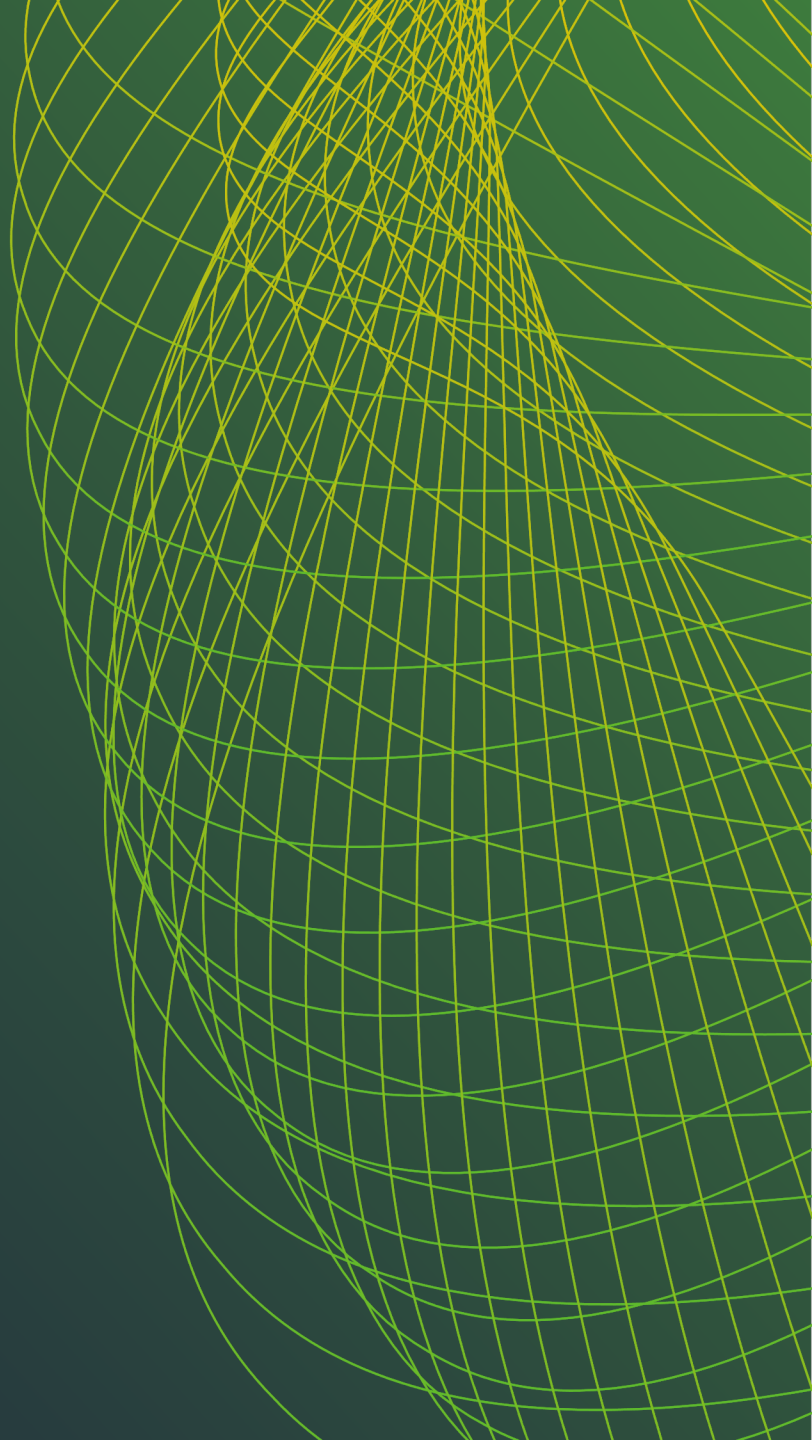
Automate and modernize your manual document workflows with AI and Serverless architecture patterns

Mojgan Ahmadi

Principal Solutions Architect
AWS

David Kheyman

Community Solutions Architect
AWS



Agenda Layout

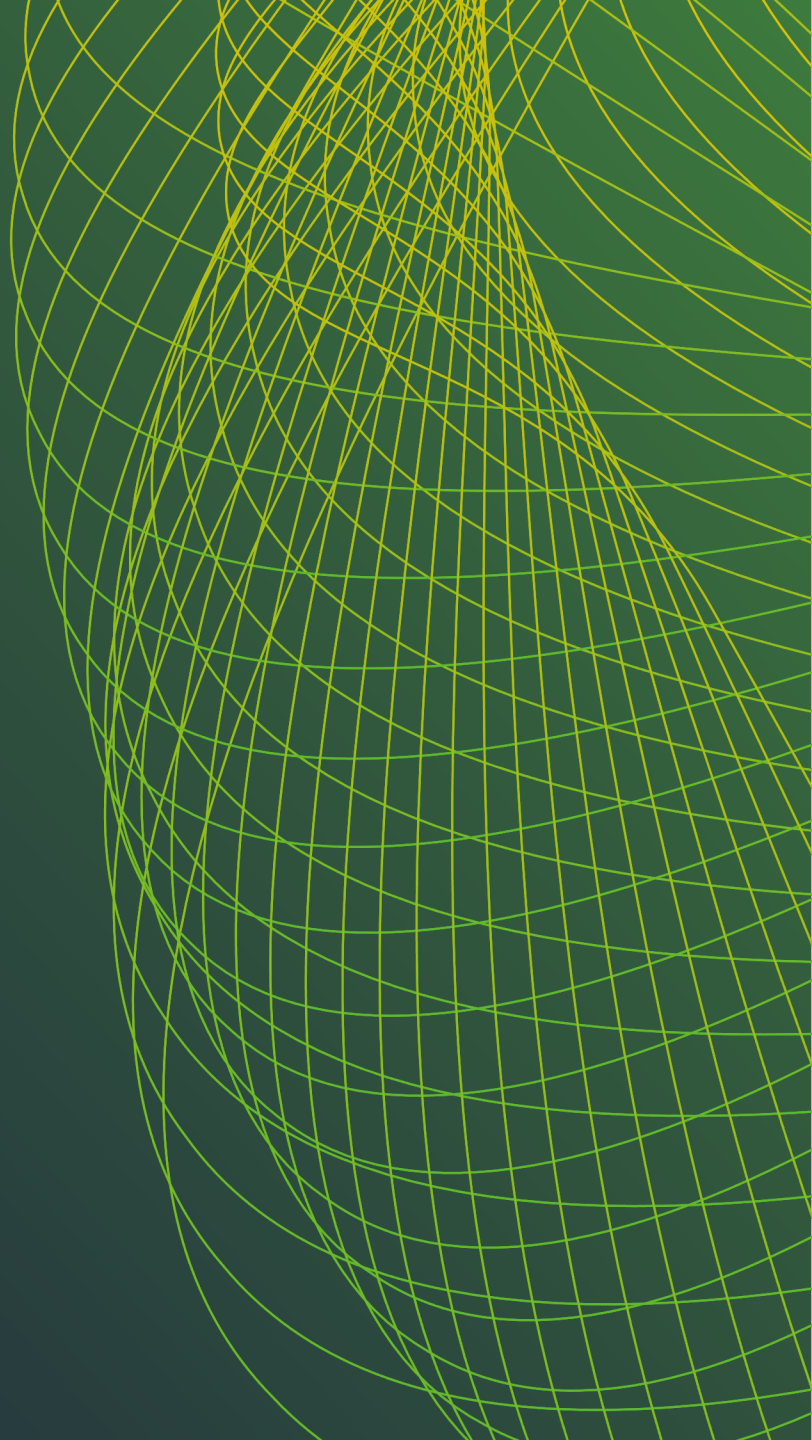
- Manual process
- Architecture of an NLP pipeline
- Open Source code

FINANCIAL SERVICES
CLOUD SYMPOSIUM

- Legacy data and its impact on application modernization
- From paper to paperless by use of AWS AI Services

aws FINANCIAL SERVICES
CLOUD SYMPOSIUM

NLP pipeline



Architecture Components

- Modular Design
- Event Driven Architecture
- Serverless

Services used

- Serverless Services


Amazon SNS


Amazon SQS


AWS Lambda

- AI Services


Amazon Textract


Amazon Comprehend

- Storage


Amazon S3

- Analytics Services


Amazon DynamoDB

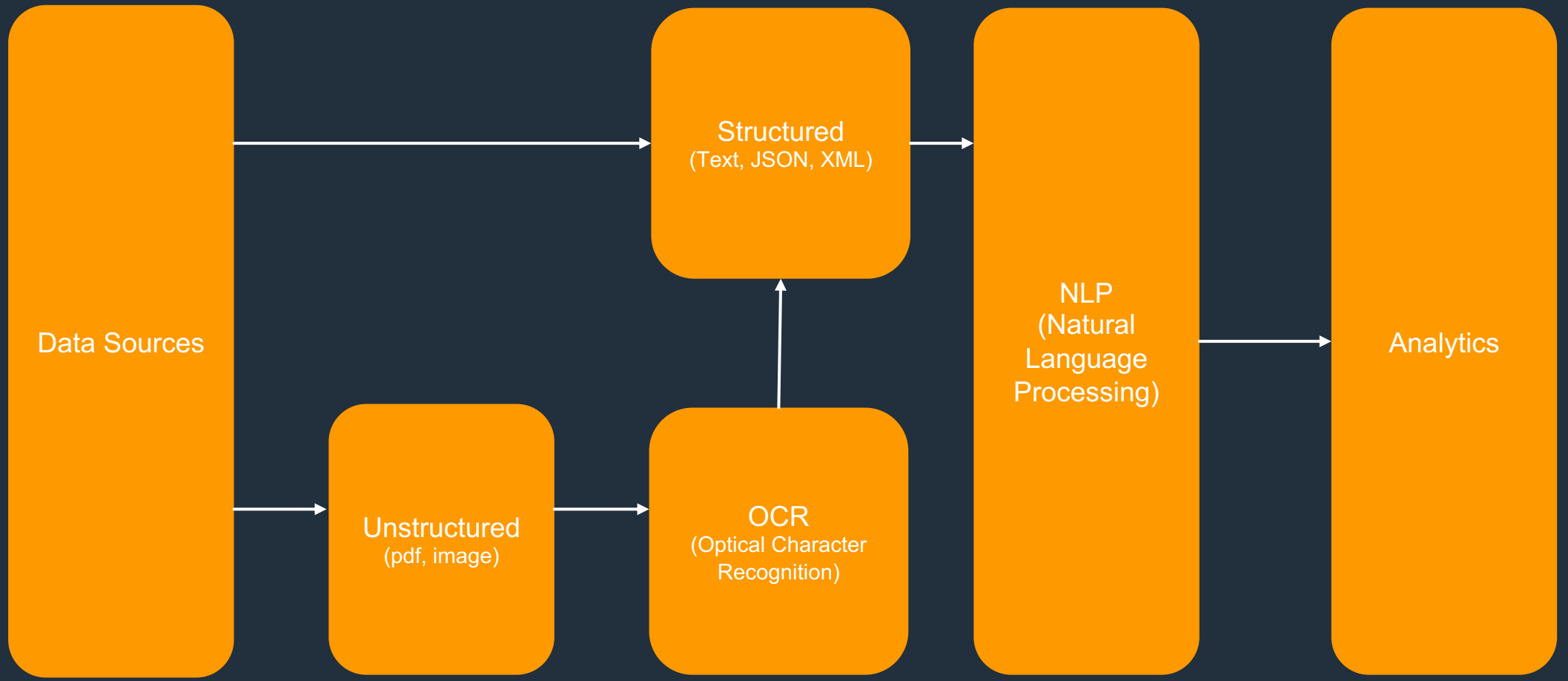

Amazon Neptune


Amazon Elasticsearch

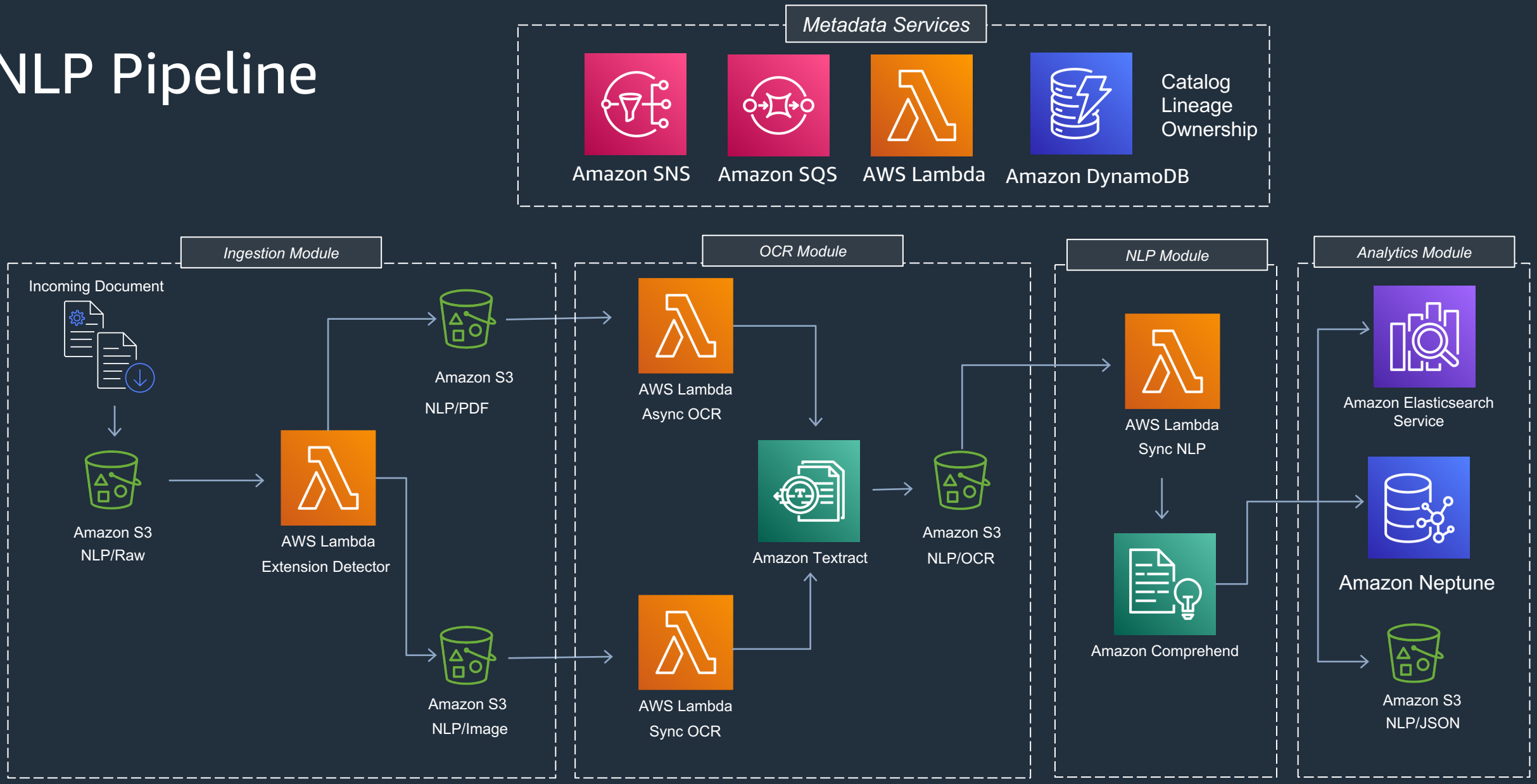
NLP Pattern

- How to design an end to end NLP pipeline
- How to incorporate data governance and lineage
- Ingestion
- OCR (optical character recognition)
- NLP (natural language processing)
- Consumption

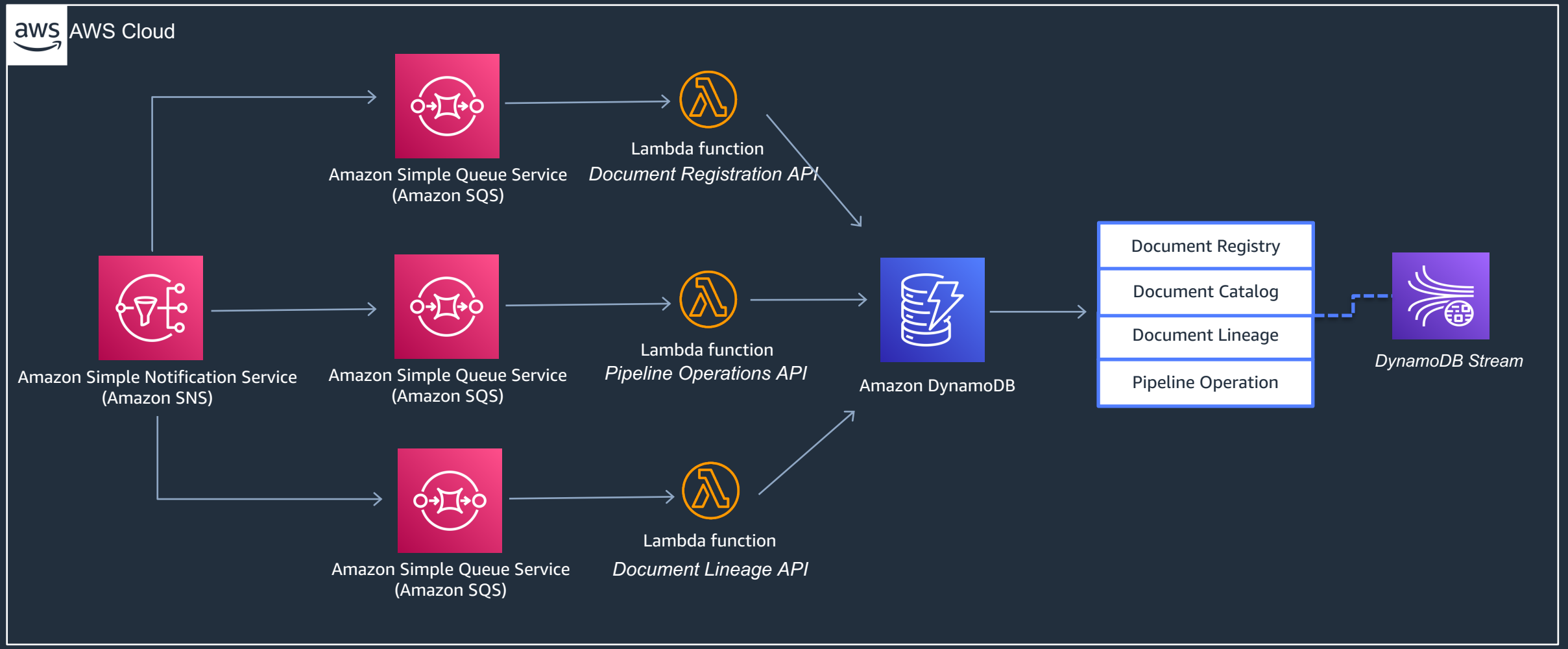
NLP – Conceptual Design



NLP Pipeline

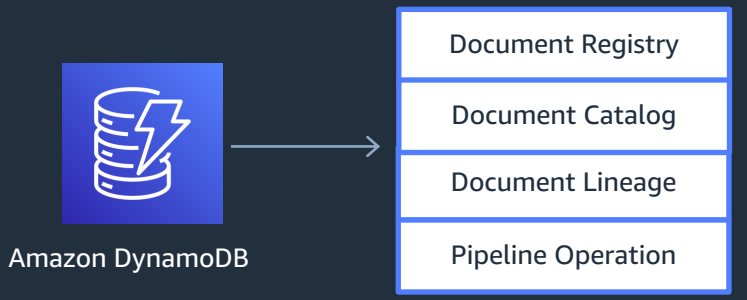


Metadata Services

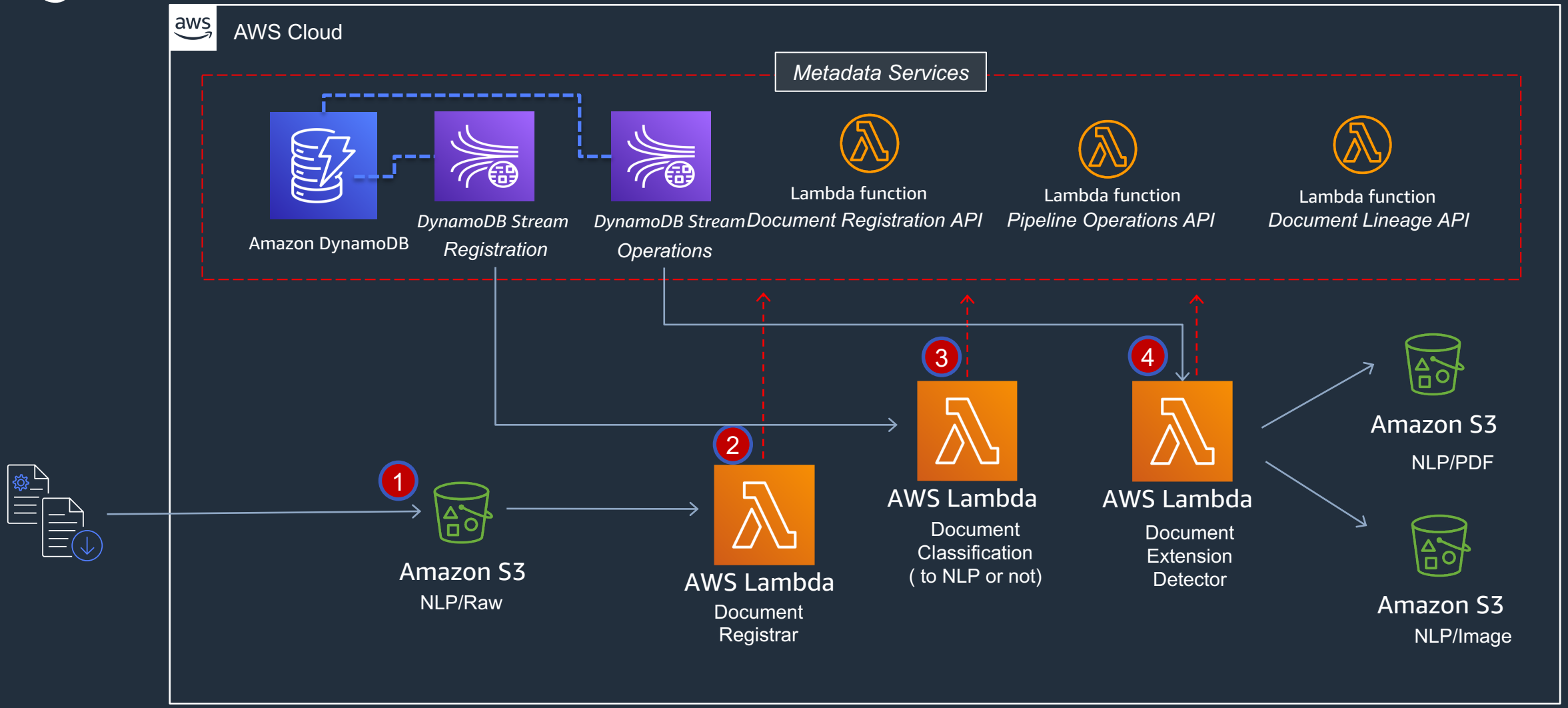


Metadata Services Table Schema

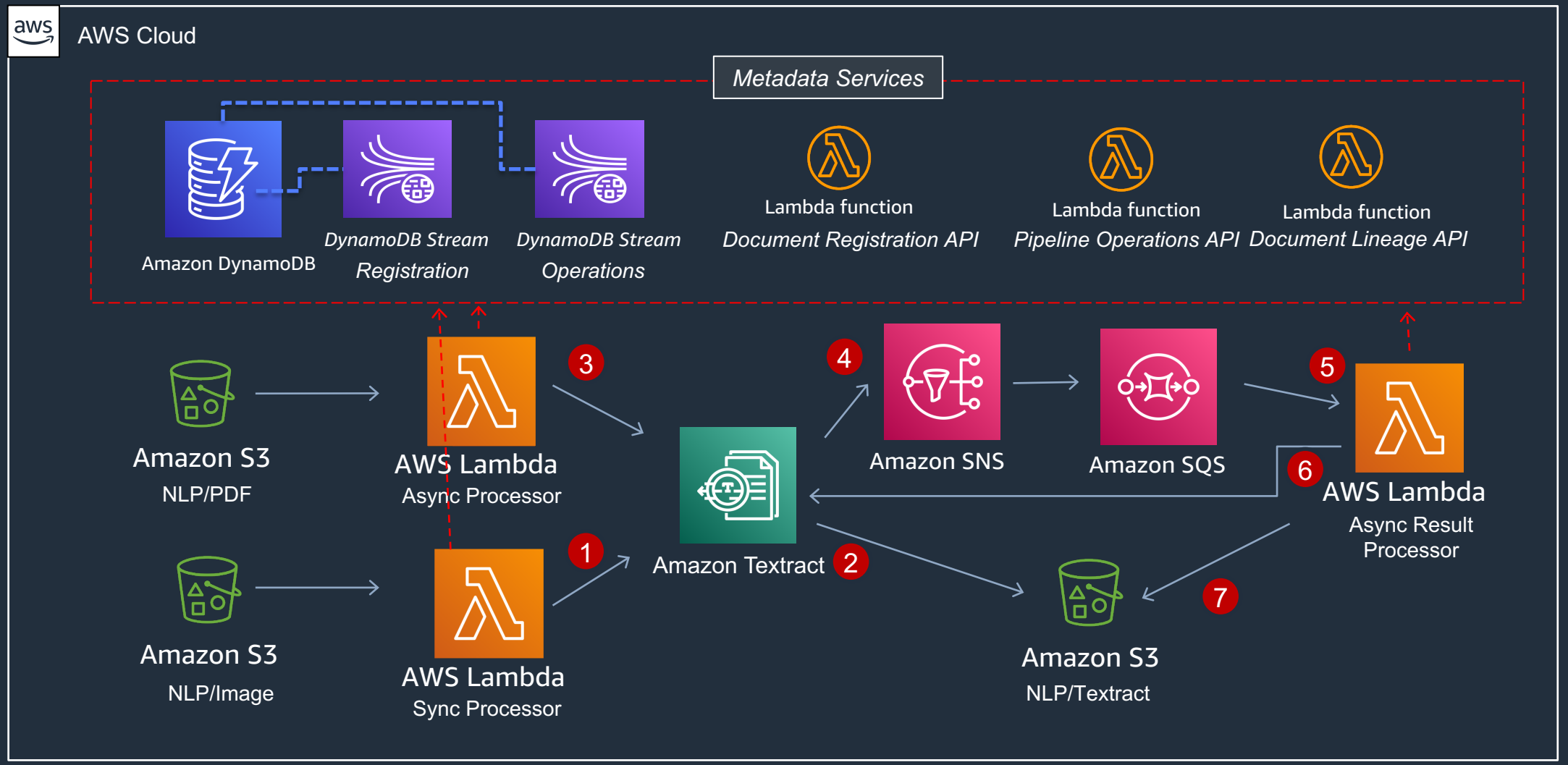
DocumentRegistry	
documentId	Unique ID generated by metadata services and assigned to this document
bucketName	S3 bucket name
documentLink	Document URI (Uniform Resource Identifier)
documentMetadata	JSON format metadata supplied by the caller
principalIAMWriter	IAM ID of the caller
timestamp	Timestamp
DocumentCatalog	
documentId	Unique ID generated by metadata services and assigned to this document
ownerName	Business user owner name
lineOfBusiness	Line of business this document belongs to
documentType	Document Type. Sampe values: 10K,10Q, CSR (Corporate Social Responsibility Report)
documentContent	Content description
ownerNotificationEmail	Owner notification email
DocumentLineage	
documentId	Unique ID generated by metadata services and assigned to this document
timestamp	Timestamp
callerId	Caller IAM ID who registered the lineage info
documentSignature	Composite Key - Unique ID assigned to this row
s3Event	S3 events - Sample Values: PUT, COPY, etc.
sourceBucketName	Source S3 bucket name
sourceFileName	Source file name
targetBucketName	Target bucket name
targetFileName	Target file name
PipelineOperation	
documentId	Unique ID generated by metadata services and assigned to this document
bucketName	S3 bucket name
documentStage	This column is updated by current process running on the document. Sample values: Comprehend, Lambda
documentStatus	Latest status on the column "documentStage"
lastUpdate	Latest timestamp when column "documentStatus" was updated
objectName	Document name
timeline	JSON document to include all operations across the pipeline



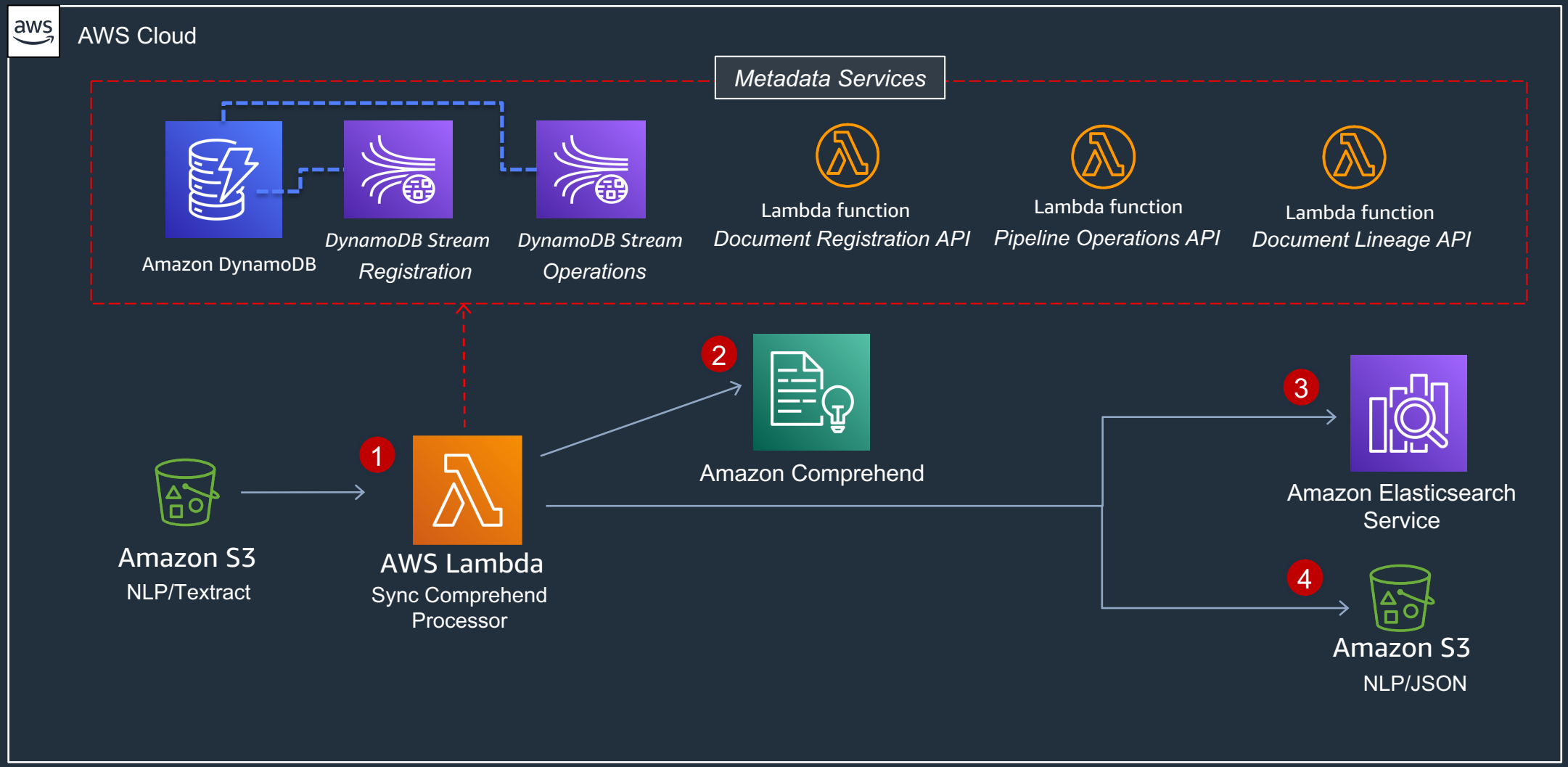
Ingestion Module



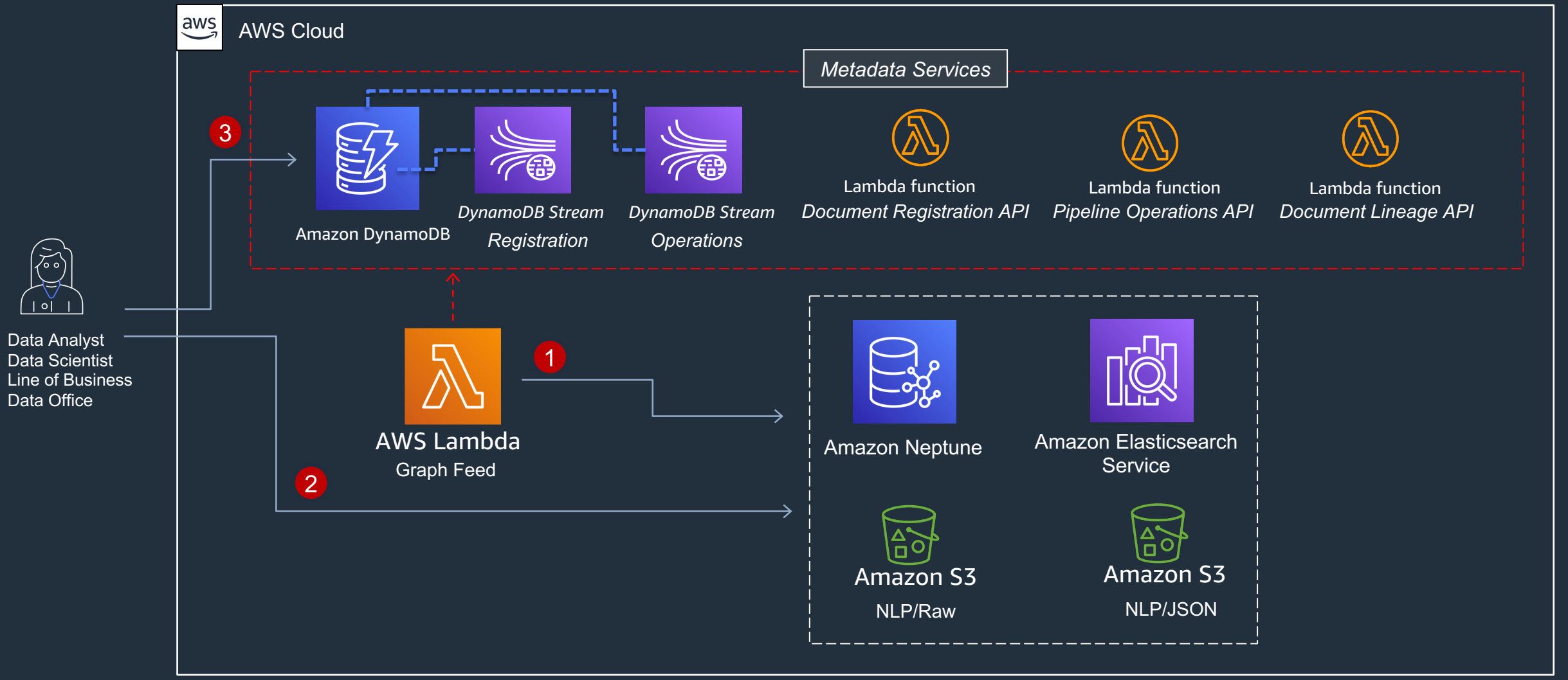
OCR Module


AWS Cloud
Metadata Services
Amazon DynamoDB
DynamoDB Stream
Registration
DynamoDB Stream
Operations
Lambda function
Document Registration API
Lambda function
Pipeline Operations API
Lambda function
Document Lineage API
Amazon S3
NLP/PDF
AWS Lambda
Async Processor
3
4
Amazon SNS
Amazon SQS
5
AWS Lambda
Async Result
Processor
6
Amazon Textract
1
2
Amazon S3
NLP/Image
AWS Lambda
Sync Processor
7
Amazon S3
NLP/Textract

NLP Module


AWS Cloud
Metadata Services
Amazon DynamoDB
DynamoDB Stream
Registration
DynamoDB Stream
Operations
Lambda function
Document Registration API
Lambda function
Pipeline Operations API
Lambda function
Document Lineage API
1
2
3
4
Amazon S3
NLP/Textract
AWS Lambda
Sync Comprehend
Processor
Amazon Comprehend
Amazon Elasticsearch
Service
Amazon S3
NLP/JSON

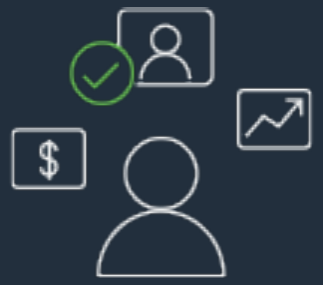
Analytics Module



FINANCIAL SERVICES
CLOUD SYMPOSIUM

Solution Walk-Through

Solution Design Principles



Separation of Duties

Solution enables data analytics, security, finance, and pipeline operations with specific components

Time Savings

Use of Infrastructure as Code (IaC) and native Automation

100% written in AWS Cloud Development Kit (CDK)

Customer Collaboration

Iterative development process based on customer needs and feedback

Solution Flexibility

Code written to fit document types for any use case

Customers can bring their own ML and analytics tooling

Downstream Applications

Solution can be easily expanded with team-specific workflows like analyzing structured tables in documents with Relational Databases

Deploying the Solution


```
git clone <repo name> pipeline

cd pipeline/code

bash build.sh && cd ../infrastructure

npm install –i

cdk bootstrap

cdk deploy --all

... and you are now ready to upload documents
to your raw bucket!
```

https://github.com/aws-samples/document-processing-pipeline-for-regulated-industries.git

Visualizing the Solution

Original Document Excerpt

Information About Our Executive Officers

Name	Age	Position
Jeffrey P. Bezos	56	President, Chief Executive Officer, and Chairman of the Board
Jeffrey M. Blackburn	50	Senior Vice President, Business Development
Andrew R. Jassy	52	CEO Amazon Web Services
Brian T. Olsavsky	56	Senior Vice President and Chief Financial Officer
Shelley L. Reynolds	55	Vice President, Worldwide Controller, and Principal Accounting Officer
Jeffrey A. Wilke	53	CEO Worldwide Consumer
David A. Zapolsky	56	Senior Vice President, General Counsel, and Secretary

Jeffrey P. Bezos. Mr. Bezos has been Chairman of the Board of Amazon.com since founding it in 1994 and Chief Executive Officer since May 1996. Mr. Bezos served as President of the Company from founding until June 1999 and again from October 2000 to the present.

Jeffrey M. Blackburn. Mr. Blackburn has served as Senior Vice President, Business Development, since April 2006.

Andrew R. Jassy. Mr. Jassy has served as CEO Amazon Web Services since April 2016, and Senior Vice President, Amazon Web Services, from April 2006 until April 2016.

Brian T. Olsavsky. Mr. Olsavsky has served as Senior Vice President and Chief Financial Officer since June 2015, Vice President, Finance for the Global Consumer Business from December 2011 to June 2015, and numerous financial leadership roles across Amazon with global responsibility since April 2002.

Shelley L. Reynolds. Ms. Reynolds has served as Vice President, Worldwide Controller, and Principal Accounting Officer since April 2007.

Jeffrey A. Wilke. Mr. Wilke has served as CEO Worldwide Consumer since April 2016, Senior Vice President, Consumer Business, from February 2012 until April 2016, and as Senior Vice President, North America Retail, from January 2007 until February 2012.

David A. Zapolsky. Mr. Zapolsky has served as Senior Vice President, General Counsel, and Secretary since May 2014, Vice President, General Counsel, and Secretary from September 2012 to May 2014, and as Vice President and Associate General Counsel for Litigation and Regulatory matters from April 2002 until September 2012.

Board of Directors

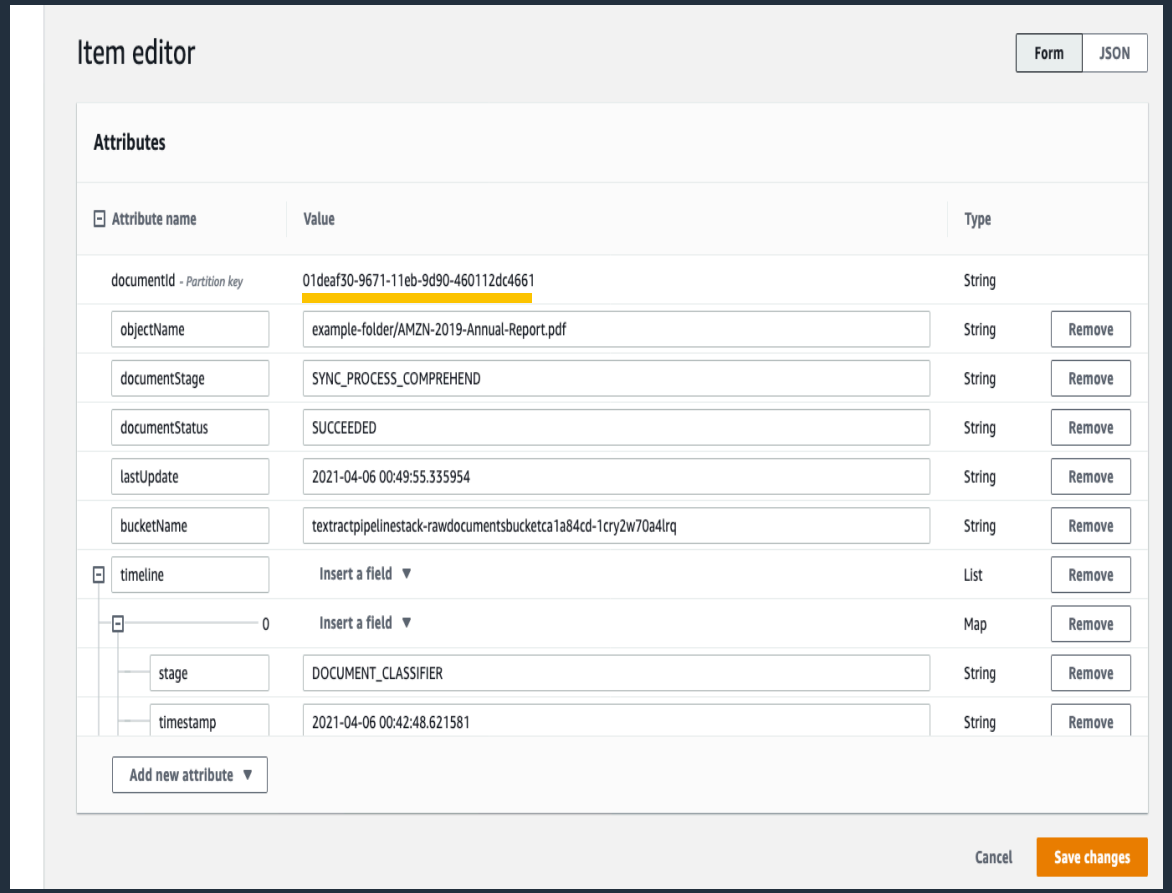
Name	Age	Position
Jeffrey P. Bezos	56	President, Chief Executive Officer, and Chairman of the Board
Rosalind G. Brewer	57	Group President, Americas and Chief Operating Officer, Starbucks Corporation
Jamie S. Gorelick	69	Partner, Wilmer Cutler Pickering Hale and Dorr LLP
Daniel P. Huttenlocher	61	Dean, MIT Schwarzman College of Computing
Judith A. McGrath	67	Senior Advisor, Astronauts Wanted * No experience necessary
Indra K. Nooyi	64	Former Chief Executive Officer, PepsiCo, Inc.
Jonathan J. Rubinstein	63	Former co-CEO, Bridgewater Associates, LP
Thomas O. Ryder	75	Retired, Former Chairman, Reader's Digest Association, Inc.
Patricia Q. Stonesifer	63	Former President and Chief Executive Officer, Martha's Table
Wendell P. Weeks	60	Chief Executive Officer, Corning Incorporated

5

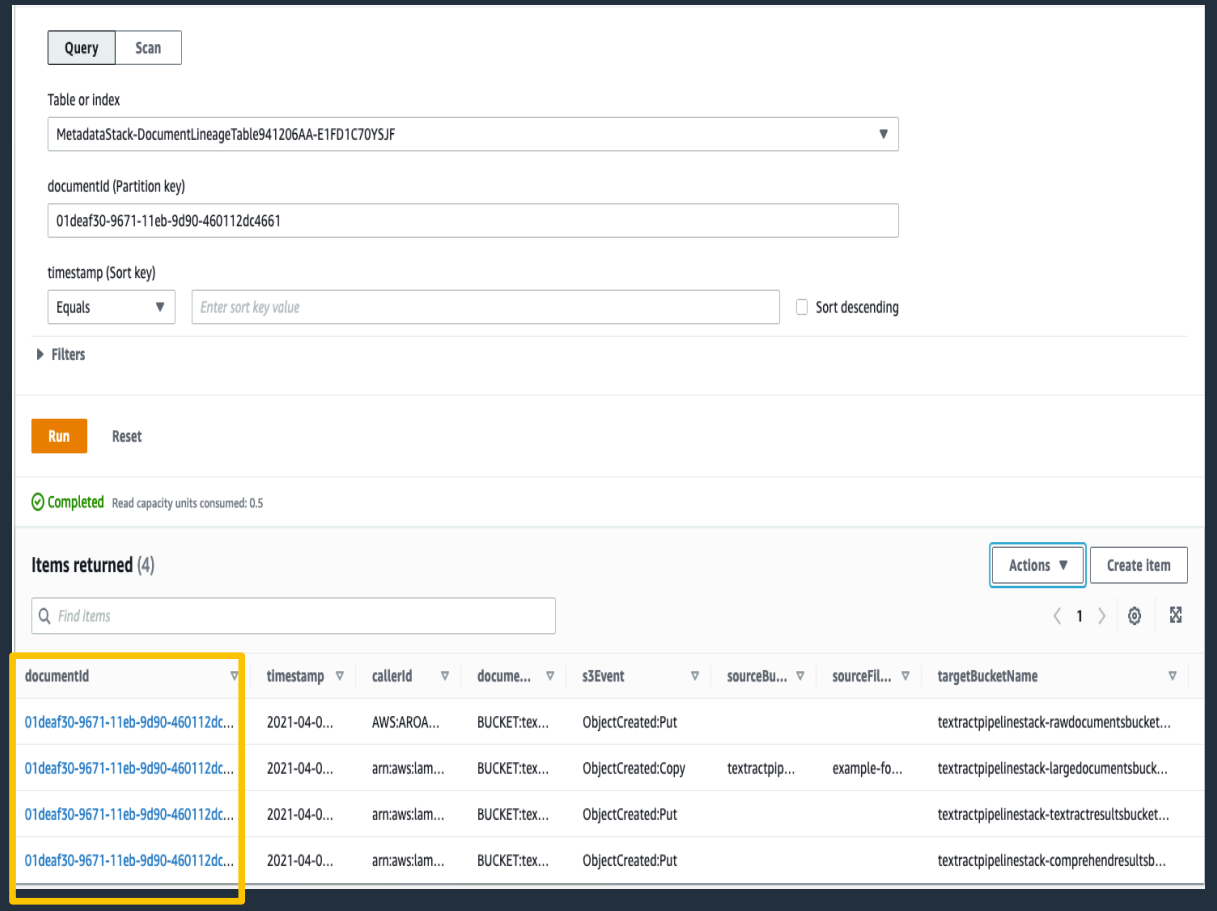
Processing Results Excerpt In Amazon Elasticsearch

Governing the Solution

Pipeline Operations Table



Data Lineage Table



FINANCIAL SERVICES
CLOUD SYMPOSIUM

Thank you!

Mojgan Ahmadi
mojgana@amazon.com

David Kheyman
kheyman@amazon.com

aws

FINANCIAL SERVICES
CLOUD SYMPOSIUM

Please complete the session survey