AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

FSI303

Unlock business value with ML document processing pipelines

Mojgan Ahmadi (she/her)
Principal Solutions Architect
Amazon Web Services

David Kheyman (he/him)
Solutions Architect
Amazon Web Services

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

Agenda

Current pain points with document processing

Breaking down the problem

Piecing together a solution

Demo

Use cases forum

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

Current pain points

HOW IS TEXT AND PDF DOCUMENT PROCESSING SUPPORTED TODAY?

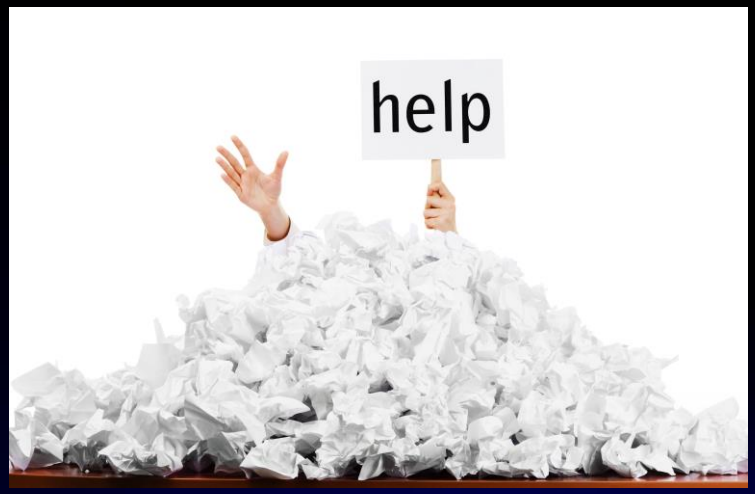
Highly manual





Lack of transparency

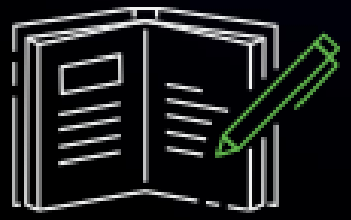
Not scalable



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

Architecture components

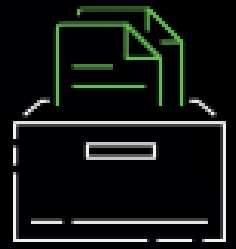
CAN YOU EXPLAIN IN SIMPLE TERMS WHAT COMPONENTS ARE INVOLVED?



Extract



Process



Store

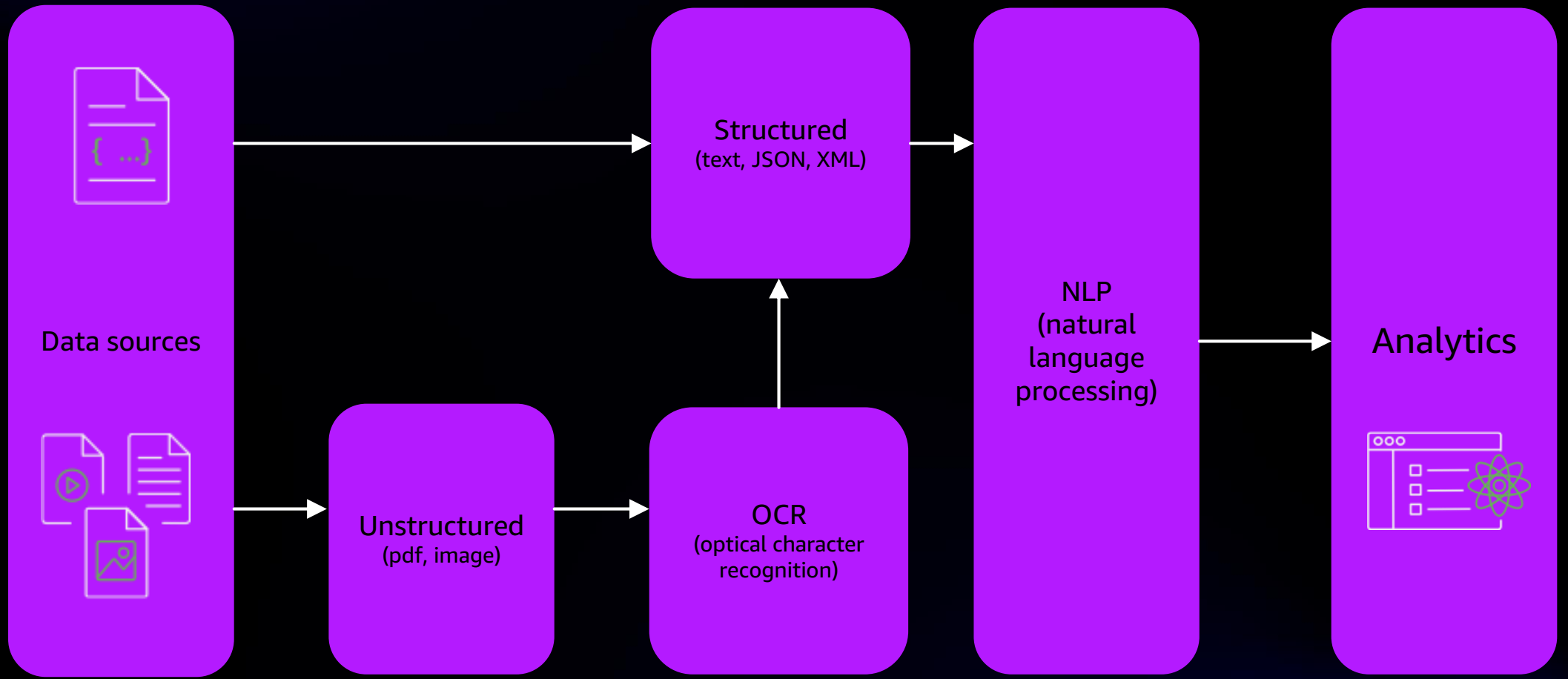


Reference

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

Logical architecture

NLP PIPELINE FRAMEWORK

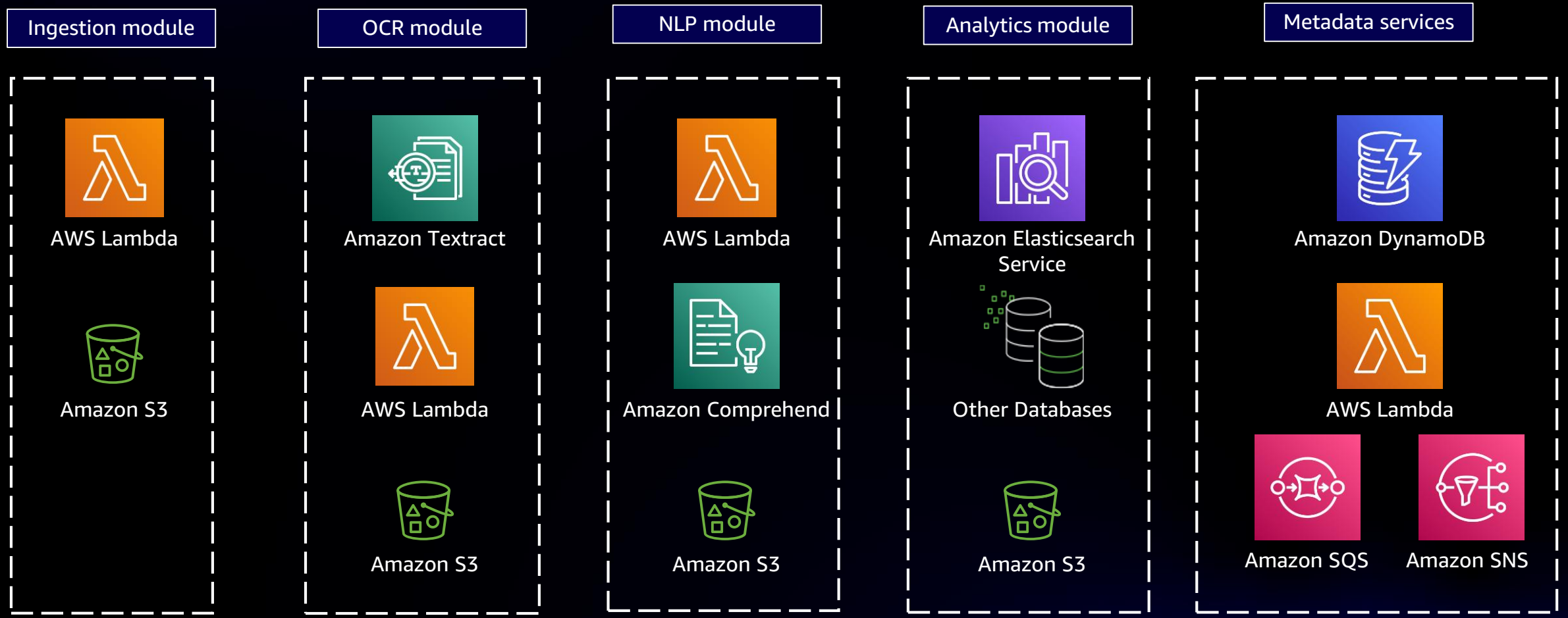


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How do your systems currently work?

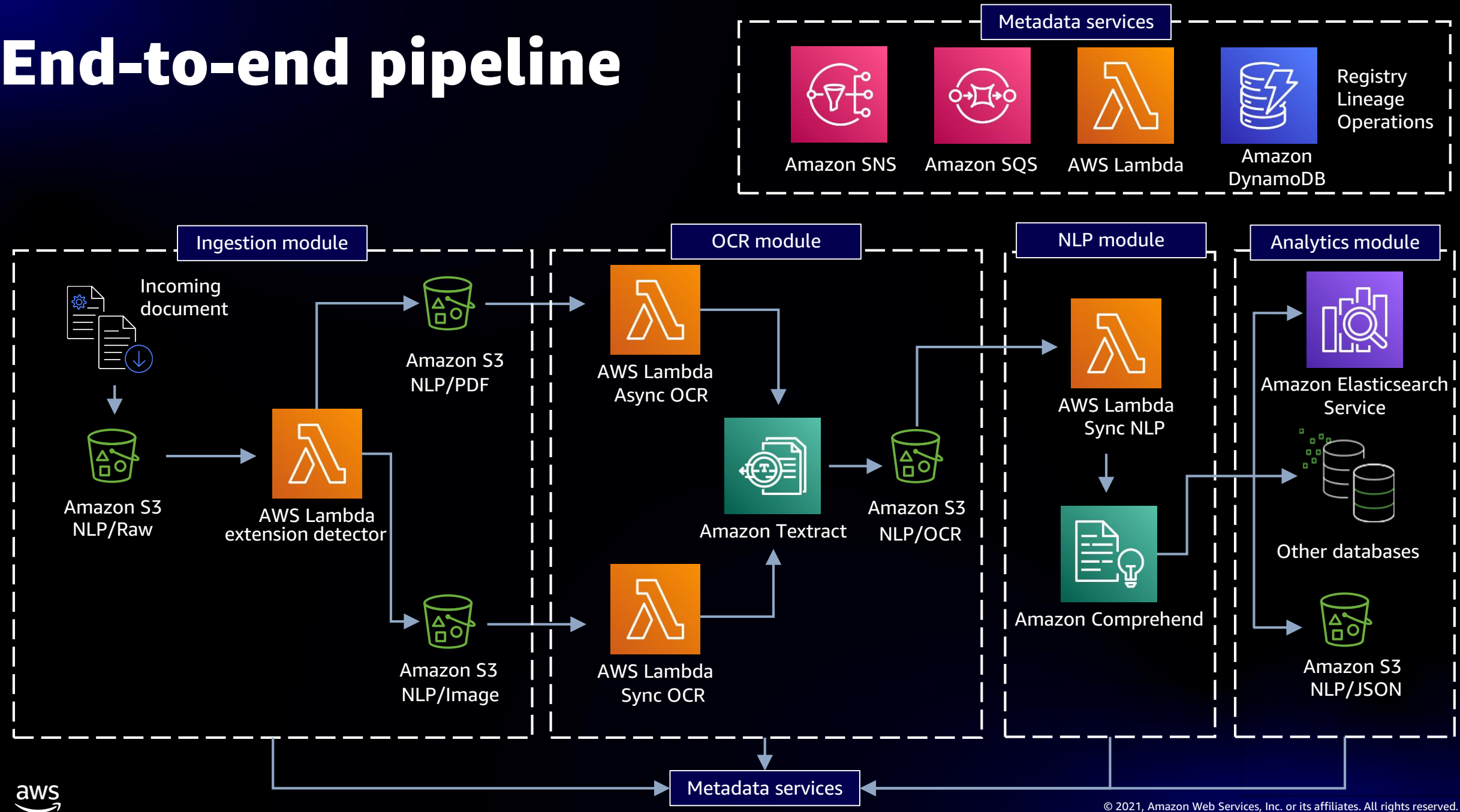
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Technology design modules



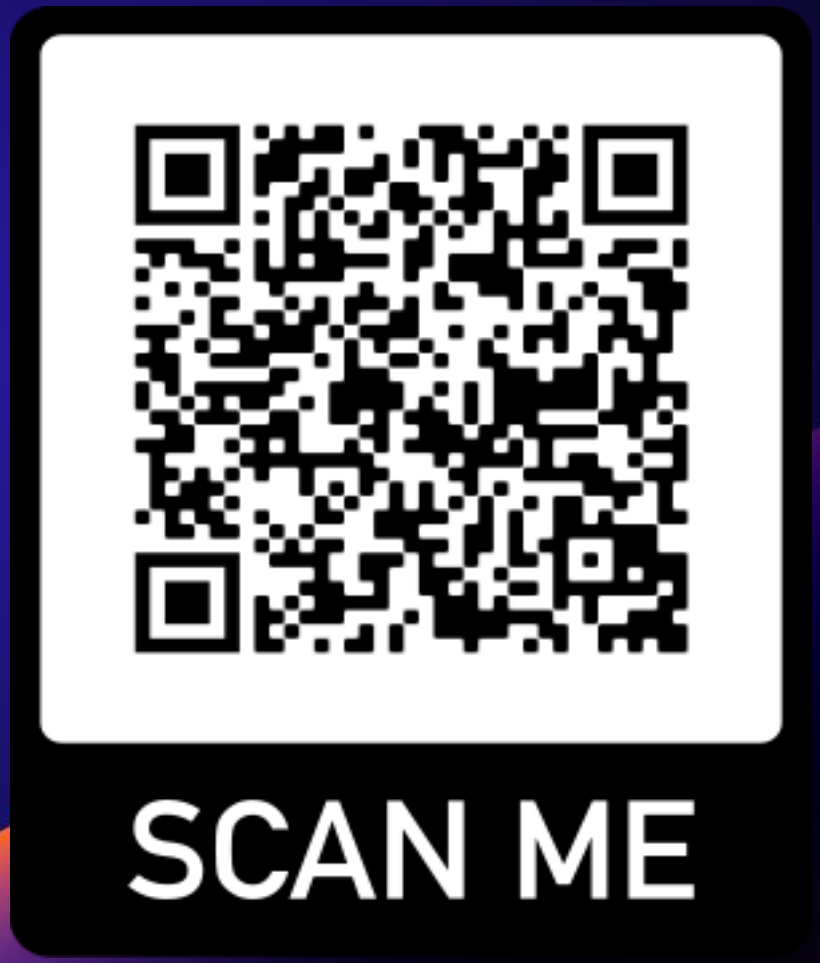
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

End-to-end pipeline



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

Use cases



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

Thank you!

Mojgan Ahmadi

mojgana@amazon.com

David Kheyman

kheyman@amazon.com
Github: dkheyman

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