Automate for Efficiency with Amazon Transcribe & Amazon Translate
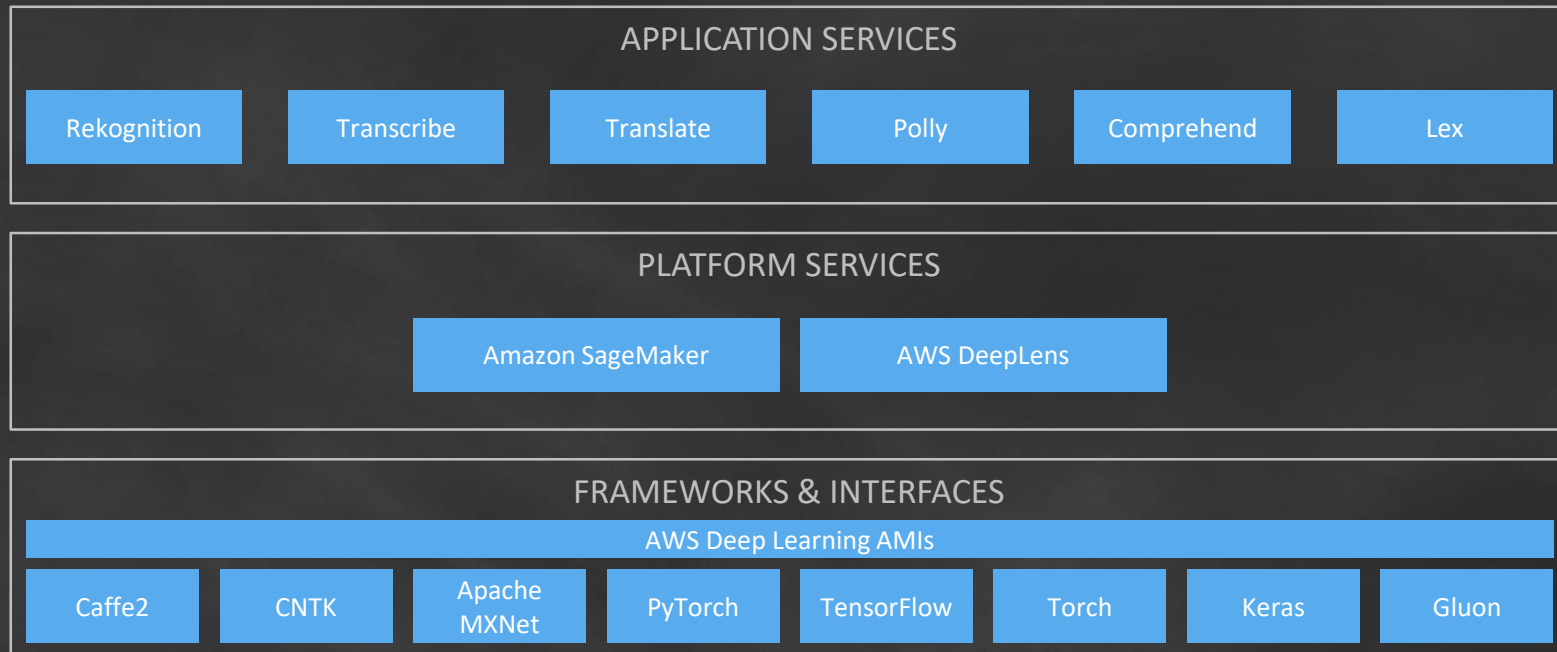
Hira, Niranjan
AWS Solutions Architect

aws | Pop-up Loft

The Amazon machine learning stack

APPLICATION SERVICES

Rekognition	Transcribe	Translate	Polly	Comprehend	Lex

PLATFORM SERVICES

Amazon SageMaker	AWS DeepLens

FRAMEWORKS & INTERFACES

AWS Deep Learning AMIs

Caffe2	CNTK	Apache MXNet	PyTorch	TensorFlow	Torch	Keras	Gluon

Amazon Transcribe

Automatic speech recognition

Multiple languages

Intelligent punctuation and formatting

Timestamp generation

Support for telephony audio

Recognize multiple speakers

Custom vocabulary

Amazon Translate

Natural and fluent language translation (based on neural networks)

Real-time translation

Batch analysis

Automatic language recognition

Low cost

Amazon Comprehend

Natural language processing

Amazon
Comprehend

Entities

Key Phrases

Language

Sentiment

Common Use Cases for Language Services

Voice of Customer Applications

Knowledge Management

Education

Customer Service/ Call Centers

Semantic Search

Accessibility

Enterprise Digital Assistant

Captioning Workflows

Information Bots

Personalization
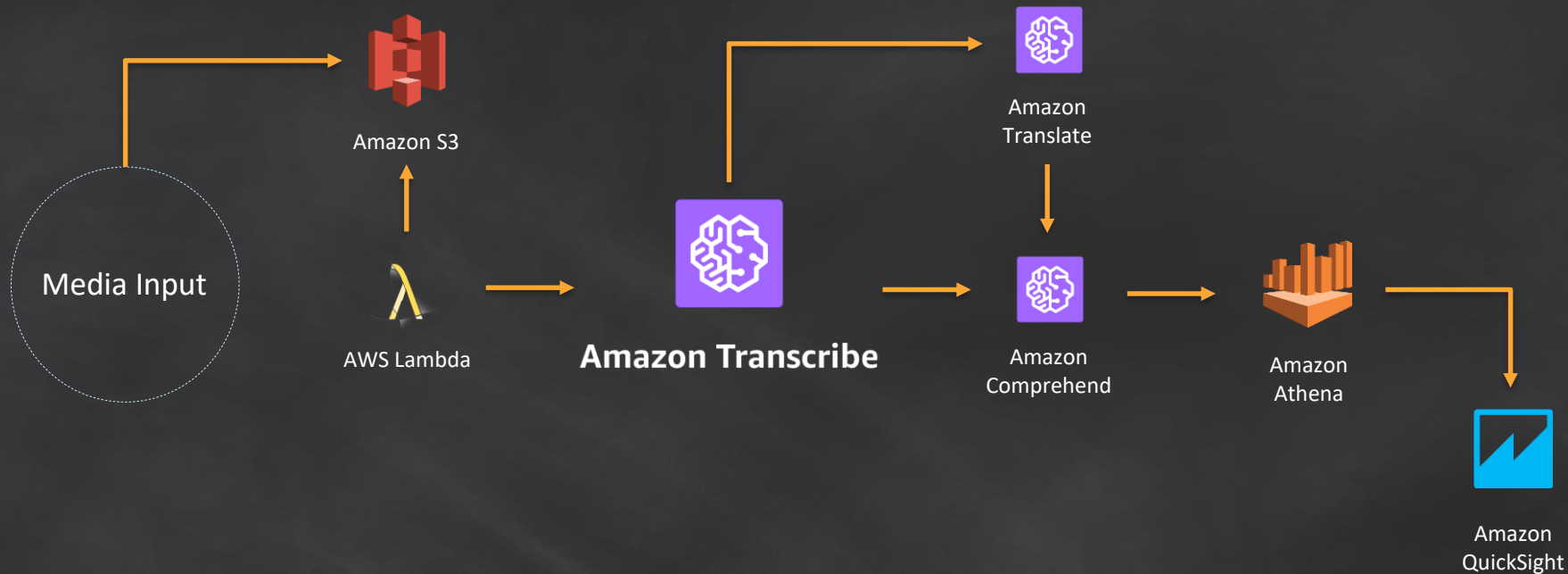
Localization

Typical pipeline

Media Input

Amazon S3

AWS Lambda

Amazon Transcribe

Amazon Translate

Amazon Comprehend

Amazon Athena

Amazon QuickSight

aws

Demo – Amazon Transcribe & Amazon Comprehend

Submit Transcribe Job → Wait for Job → Get Transcript → Detect Entities → Write HTML Wrapper
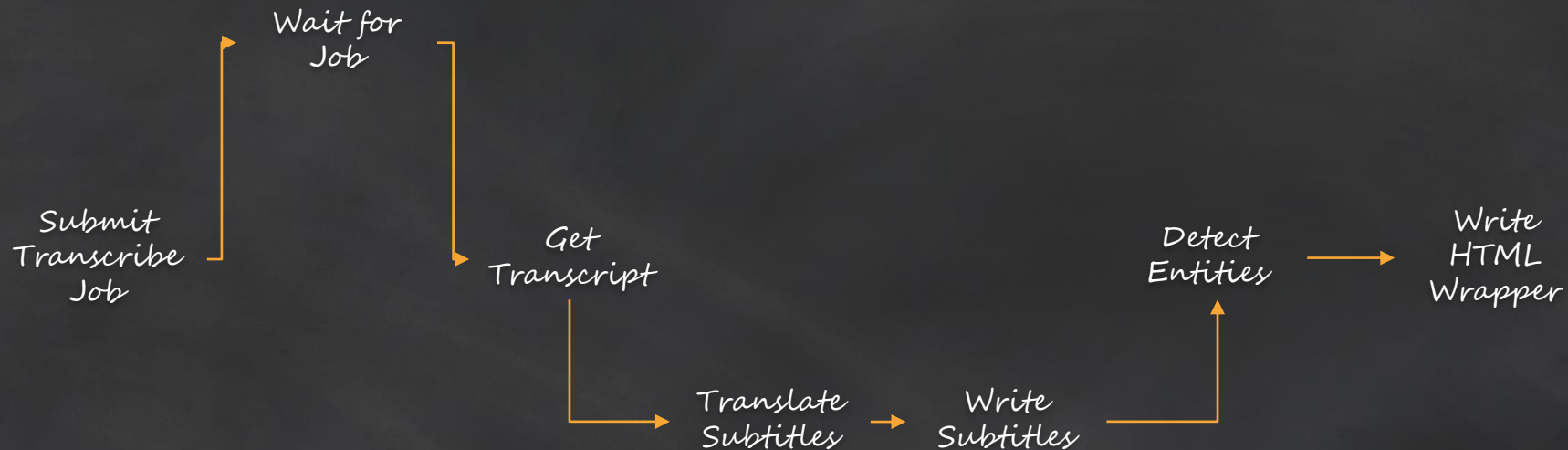
▶ 0:00 / 15:20 🔊

[00:00:00,000] You're listening to the six o'clock news podcast from BBC
[00:00:03,600] radio four labor
[00:00:06,870] has said it backs remaining in a customs union with you
[00:00:10,130] after Brexit, in a policy shift ahead of major speeches
[00:00:13,280] this week by both jeremy corbyn on theresa may,
[00:00:17,030] the winter olympics in South Korea have ended after sixteen
[00:00:20,130] days of sporting and political drama.
[00:00:23,530] Clashes between Syrian government and rebel forces have intensified
[00:00:27,120] around eastern ghouta, a day after the UN demanded a cease
[00:00:30,410] fire,
[00:00:31,880] China's communist party has announced plans that would allow president
[00:00:35,080] xi jinping to stay in power indefinitely.
[00:00:40,370] The shadow Brexit secretary secure starmer has confirmed
[00:00:43,450] that labor wants the UK to remain in a customs union
[00:00:46,470] with EU after Brexit. The labour leader jeremy
[00:00:49,750] corbyn is to set out the party's new position in a speech tomorrow.
[00:00:53,590] The government is committed to leaving both the single market and the customs
[00:00:57,020] union, but a number of conservative mps have said they're ready
[00:01:00,110] to vote with labor to keep Britain in a customs union.
[00:01:03,840] The international trade secretary, liam fox, has urged them
[00:01:06,950] to keep an open mind, suggesting that the prime minister would
[00:01:09,960] deal with a lot of their concerns when she makes a speech later
[00:01:13,250] this week. Here's, our political correspondent, emma body,
[00:01:17,240] after months of uncertainty about labor's position,
[00:01:20,320] the party is now moving decisively away from a so called
[00:01:23,570] hard Brexit speaking on the andrew marr show circus
[00:01:27,040] starmer said jeremy corbyn would announce tomorrow that
[00:01:30,120] a labour government's policy would be to negotiate a new
[00:01:33,150] customs union with you we've long championed
[00:01:36,590] being in a customs union with the you and the benefits of that
[00:01:40,080] obviously it's the only way realistically to get tire free access
[00:01:43,400] it's really important for our manufacturing base on dh nobody
[00:01:46,800] can ask the question how you keep your commitment to know how border in

Entity	Type	Max Confidence	Count
Friday	DATE	1	2
Last night	DATE	0.98	1
Saturday	DATE	1	1
few months after	DATE	0.99	1
last week	DATE	1	1
late last year	DATE	0.68	1
mid week	DATE	1	1
next month	DATE	0.88	1
past week	DATE	1	2
six o'clock	DATE	1	2
this week	DATE	0.96	2
today	DATE	1	1
tomorrow	DATE	1	2
yesterday morning	DATE	1	1
Brexit	EVENT	1	14
Britain	LOCATION	1	2
British	LOCATION	1	1
China	LOCATION	1	8
Europe	LOCATION	1	1
European	LOCATION	1	1
French	LOCATION	1	3
German	LOCATION	0.97	2
Iran	LOCATION	0.92	1
Iraq	LOCATION	1	2
North Korea	LOCATION	1	2
North Korean	LOCATION	0.9	2
North korean	LOCATION	0.99	1
Northern Ireland	LOCATION	1	2
South Korea	LOCATION	1	3
South Korean	LOCATION	1	1

Amazon Translate

```
1. Translate.translateText({
2.     "Text": 'Hello world',
3.     "SourceLanguageCode": 'en',
4.     "TargetLanguageCode": 'es'
5.     }, function(err, data) { ... });
```

aws

Demo – Mix in Amazon Translate

Wait for Job

Submit Transcribe Job

Get Transcript

Translate Subtitles

Write Subtitles

Detect Entities

Write HTML Wrapper

BBC

Estas son las noticias de la BBC y estos otros titulares. Los políticos de los Estados Unidos en

[00:00:05,410] This is the BBC news and these other headlines. Politicians in the
[00:00:08,490] United States in Europe are summoning executives from
[00:00:11,590] facebook to answer questions about whether personal data
[00:00:14,710] was misused to manipulate it. The recent elections,
[00:00:18,150] facebookers said it was deceived by a British company, cambridge
[00:00:21,350] analytica, that worked for president. Trump's election campaign,
[00:00:24,450] among others. Emergency services in texas, a and
[00:00:27,610] essentially device that went off in the city of austin, was not
[00:00:30,810] related to a series of explosions this month. However,
[00:00:33,870] the FBI has confirmed that two packages found it separates
[00:00:37,070] a fedex delivery. Officers in the area on Tuesday
[00:00:40,140] are connected to the earlier attacks.
[00:00:43,450] A BBC investigation has found that girls from the and mars
[00:00:46,500] arranger muslim minority who fled violence in their home
[00:00:49,680] country on being trafficked into prostitution in
[00:00:52,700] Bangladesh, hundreds of thousands of revenge and now live in huge
[00:00:55,890] refugee camps where they're vulnerable to sexual exploitation.

Entity	Type	Max Confidence	Count
Tuesday	DATE	1	1
this month	DATE	0.97	1
Bangladesh	LOCATION	1	1
British	LOCATION	0.58	1
Europe	LOCATION	1	1
United States	LOCATION	0.99	1
austin	LOCATION	0.95	1
texas	LOCATION	0.99	1
BBC	ORGANIZATION	0.97	2
FBI	ORGANIZATION	1	1
cambridge analytica	ORGANIZATION	0.69	1
facebook	ORGANIZATION	0.99	1
fedex	ORGANIZATION	0.75	1
muslim	ORGANIZATION	0.59	1
Trump	PERSON	1	1

Index	Avg. Confidence	Terms	Unsure
00:00:05,410	0.88	17	2
00:00:10,440	0.91	25	

What does the HTML look like?

```
1.  <video controls id="vplayr">
2.      <source src="....mp4" type="video/mp4">
3.      <track kind ="subtitles" label="English" srclang="en-us"
    src="..._en.vtt">
4.      <track kind ="subtitles" label="Spanish" srclang="es"
    src="..._es.vtt">
5.      This browser does not seem to support the video element.
6.  </video>
```

aws

What is WebVTT?

- Web Video Text Tracks format
- https://www.w3.org/TR/webvtt1/
- "Borrows" from SRT (SubRip Subtitles)

```
1.  2
2.  00:00:08.490 --> 00:00:11.000
3.   Europa están convocando a ejecutivos de
4.   facebook
5.
6.  3
7.  00:00:11.590 --> 00:00:14.000
8.   para responder preguntas sobre si los
9.   datos
```

aws

aws | Pop-up Loft

Everything and Anything Startups Need to Get Started on AWS

aws.amazon.com/activate