



7.25.17 | Microsoft Technology Center | 11 Times Square, NYC



Microsoft

Media & Entertainment Day

Data, Cloud, AI, Security

Produced by **MESA**

CELEBRATE
GREAT
TASTE

Etsy

Video Analytics: Understanding What's In Your Content

Martin Wahl (mawahl@microsoft.com)
Principal Program Manager
Azure Media Services
[Twitter@redsoxnutt](https://twitter.com/redsoxnutt)



Blink of an Eye: The Digital Media Transformation

67 Percent of OTT Companies Believe Virtual Reality is a Lasting Technology

Global IP Video Will Be 82 Percent of All Online Traffic by 2020

Beyond digital: Empowered consumers seek out tailored, inspiring content experiences that transcend platforms and can be shared – Pwc

Eighty per cent of Millennials subscribe to at least one OTT service

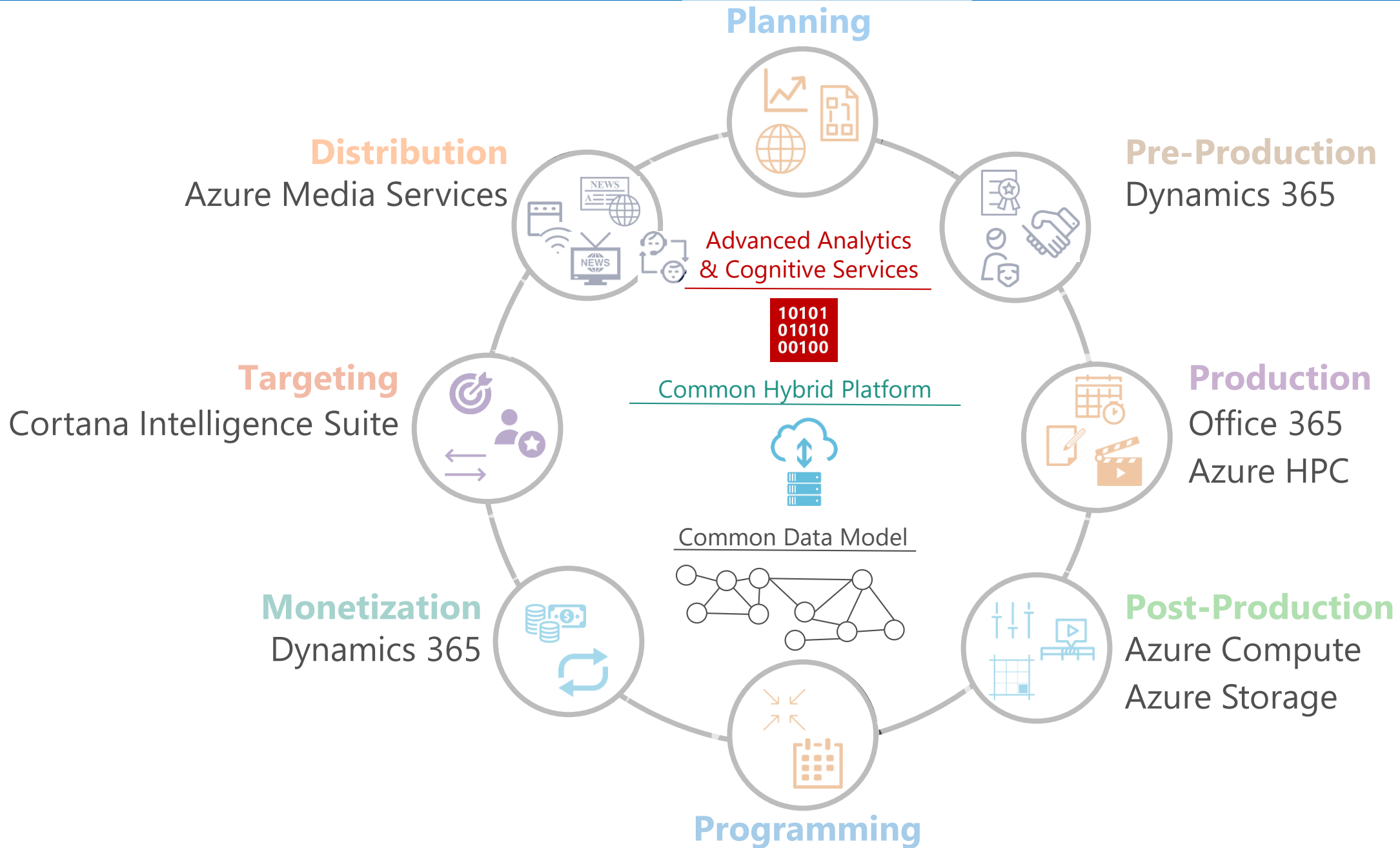
Is it really TV if it is never broadcast – growth in viewing is all digital!

By 2020 internet video surveillance traffic will be more than 5 Exabytes per month

Amount of VOD Traffic in 2020 will be equivalent to 7.2 DVDs per month

For Millennials, the smartphone is the most popular device for watching online video and gaining ground among the broader population.

Microsoft Azure: The Intelligent Cloud for Media



Azure Media Services

Azure Media Services is an extensible, multi-tenant platform that enables end-to-end video workflows in the Azure public cloud.



Primary Azure Media Scenarios

Cloud based Video Streaming

Video On-Demand Portals

- Sports
- Entertainment
- Education
- News

Live Broadcast & Events

- Sports Broadcasters
- News Channels
- Music Concerts
- Awards Ceremonies
- Town Halls / Conferences

Digital Marketing / Advertising

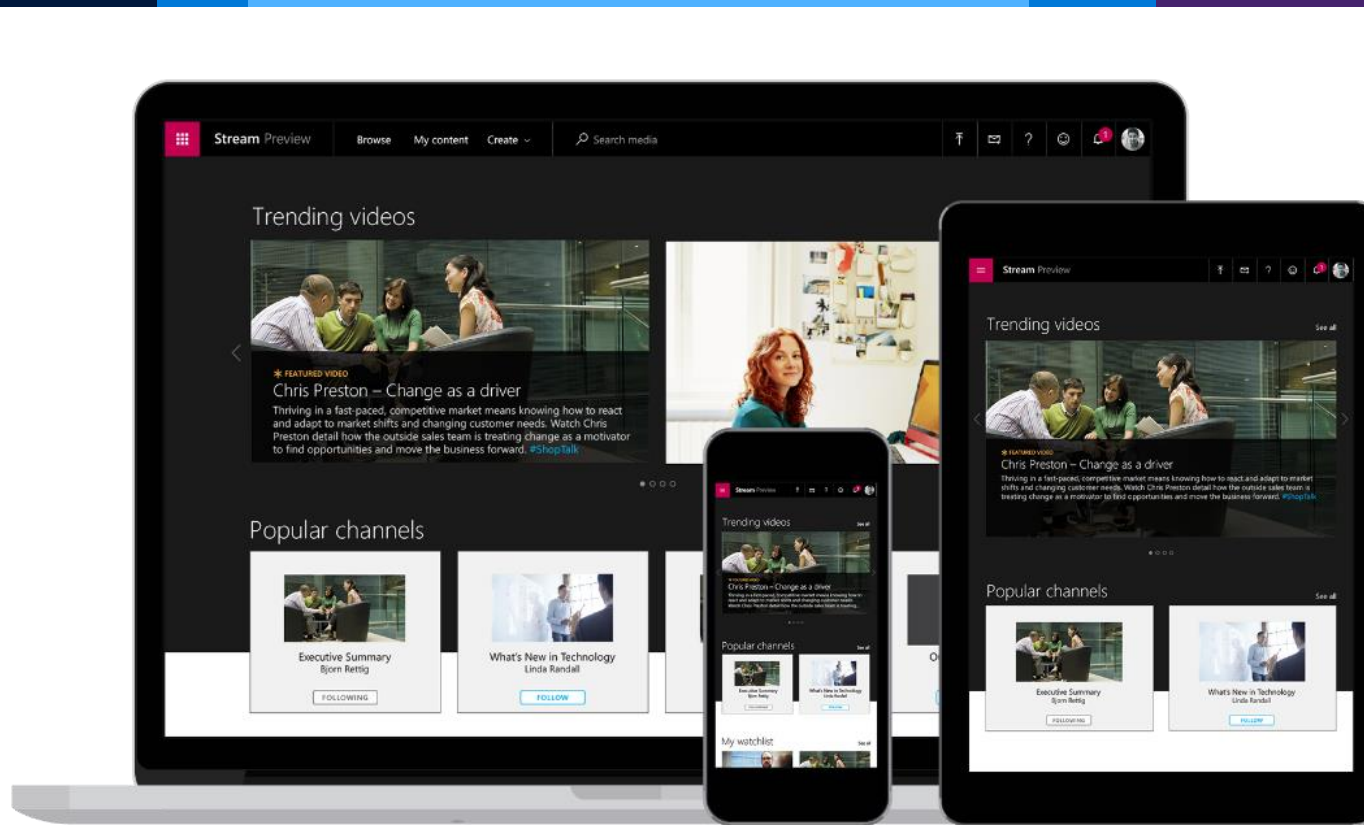
- Retail
- Finance / Banking
- Brand Awareness
- Sales Support

Enterprise / Collaborative

Corporate

CDN Use Cases

- Websites



Microsoft Stream

And the Category is...

Video AI

Imagine if you could...

- Automatically create all of the metadata (spoken text, written text, faces, places, objects) for any video, including from your archives
- Relate your advertisements dynamically to the content of the program for increased viewer relevance, value, and response
- Collaboratively upload & edit videos in real-time and through the Cloud to speed up time to air for new stories
- Enable your viewers to take advantage of deep search capabilities – search by Name, Search by Keyword, Search by Shot Location
- Distribute playable clips of live video content internally or externally without long download or encode times
- Dynamically change your content in response to audiences reaction – Remember “Choose Your Own Adventure” novels – why not do the time for modern video audiences?
- Create automated summaries or highlight reels of their video content based on scene detection, specific people, and motion within the video

Video AI is the key to solving these challenges....

- Improve Content Discoverability
- Increase Content Value
- Personalize the Viewing Experience
- Uncover Hidden Content Insights
- Increase Revenue / Drive Viewing
 - Targeted Advertising
 - Predictive modeling & recommendations
- Reduce Manual Labor



Microsoft Cognitive Services

Give your apps a human side



Vision

From faces to feelings, allow your apps to understand images and video



Speech

Hear and speak to your users by filtering noise, identifying speakers, and understanding intent



Language

Process text and learn how to recognize what users want



Knowledge

Tap into rich knowledge amassed from the web, academia, or your own data



Search

Access billions of web pages, images, videos, and news with the power of Bing APIs

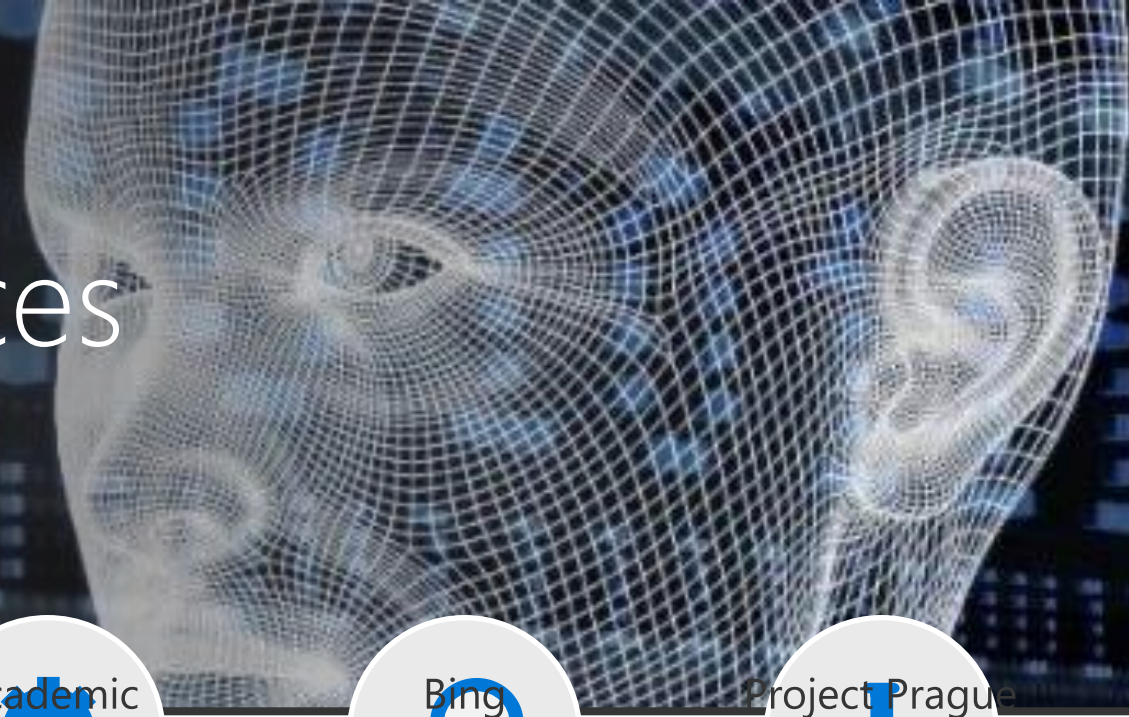


Labs

An early look at emerging Cognitive Services technologies: discover, try & give feedback on new technologies before general availability

Microsoft Cognitive Services

Give your apps a human side



Computer Vision
Content Moderator
Emotion
Vision
Face
Video
Video Indexer
Custom Vision
Service

Bing Speech
Custom Speech
Service
Speech
Recognition

Bing Spell Checker
Language
Understanding
Language
Linguistic Analysis
Text Analytics
Translator Text
& Speech
Web Language
Model

Academic
Knowledge
Entity Linking
Knowledge
Exploration
Recommendations
QnA Maker
Custom Decision
Service

Bing
Autosuggest
Search
Bing Image
Search
Bing News
Search
Bing Video
Search
Bing Web Search
Bing Custom
Search

Project Prague
(gesture)
Project Cuzco (events)
Labs
Project Limerburg
(routing)
Project Nanjing
(isochrones)
Project Abu Dhabi
(distance matrix)
Project Wollongong
(location)

Microsoft's Video AI Capabilities

Audio transcript

Convert audio to text based on acoustic language models

Face detection

Find when each face appears in the video

Video stabilization

Create smooth videos from videos captured by moving camera

Video OCR

Extract text that appears in videos as overlay, slides or background

Face redaction

Detect faces and choose which ones you want to redact

Motion detection

Detect when motion has occurred in videos

Face emotion detection

Recognize the emotion of a person or crowd based on facial expressions

Video summaries

Create summaries of long videos to enable quick previews

Content moderation

Detect and prevent explicitly visual content

Object detection

Detect objects based on a pre-defined object model

Bringing it all together....

Video Indexer

Video Indexer | Bundling Features Together

Widgets

Search

Audio transcript

Convert audio to text based on acoustic language models

Face detection

Find when each face appears in the video

Visual text recognition

Extract text that appears in videos as overlay, slides or background

Labeling

Tag objects such as cat, table, car, ball etc. when they appear

Translation

Translate audio transcript to another language

Content moderation

Detect and prevent explicitly visual and textual content

Face identification

Identify who is the person that appears in the video

Speaker indexing

Map and understand who spoke what and when

Keywords extraction

Find out the keywords discussed in the video

Sentiment analysis

Understand the level of positive vs. negative spoken or written words

Scene Detection

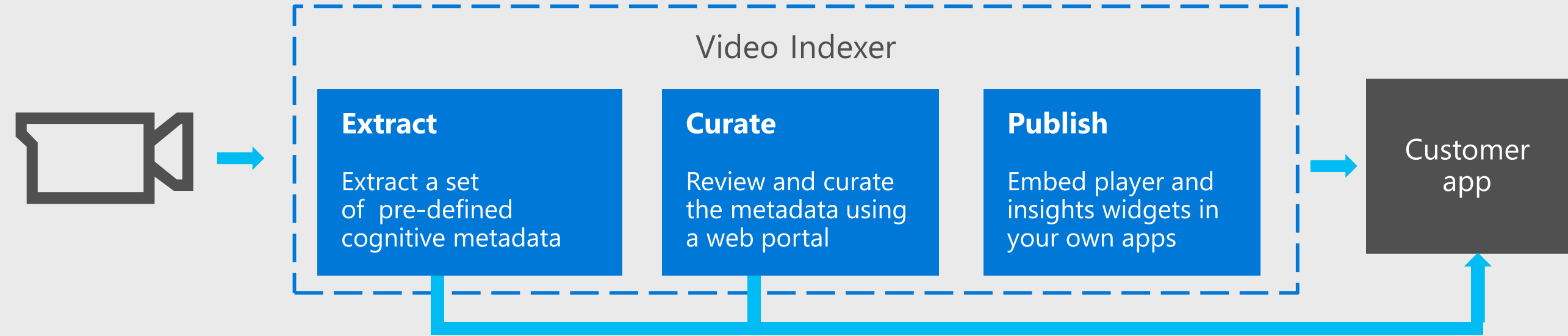
Detect when a scene changes based on visual analysis

Keyframe extraction

Automatically detect keyframes in the video

Scalable & Efficient Orchestration

Video Indexer



Video Indexer extracted metadata snippet

```
"faces": [  
  {  
    "id": 11775,  
    "shortId": "63c6d532ff",  
    "name": "Scott Hanselman",  
    "description": null,  
    "title": null,  
    "thumbnailUrl": "/api/Thumbnail/63c6d532ff/616468f0-1636-4efa-94e7-262f2e575059",  
    "thumbnailFullUrl": "https://www.videobreakdown.com/api/Thumbnail/63c6d532ff/616468f0-1636-4efa-94e7-262f2e575059",  
    "appearances": [  
      {  
        "startTime": "00:00:08.1700000",  
        "endTime": "00:00:25.3920000",  
        "startSeconds": 8.2,  
        "endSeconds": 25.4  
      },  
      {  
        "startTime": "00:12:57.6820000",  
        "endTime": "00:13:07.0580000",  
        "startSeconds": 777.7,  
        "endSeconds": 787.1  
      }  
    ],  
  ],  
]
```


Video Indexer extracted metadata snippet

```
"topics": [  
  {  
    "name": "days of history",  
    "appearances": [  
      {  
        "startTime": "00:04:42.1100000",  
        "endTime": "00:00:00",  
        "startSeconds": 282.1,  
        "endSeconds": 0.0  
      },  
      {  
        "startTime": "00:12:07.6200000",  
        "endTime": "00:00:00",  
        "startSeconds": 727.6,  
        "endSeconds": 0.0  
      }  
    ],  
    "isTranscript": true  
  }  
]
```

```
"sentiments": [  
  {  
    "sentimentKey": "Negative",  
    "appearances": [  
      {  
        "startTime": "00:00:00",  
        "endTime": "00:00:03.5000000",  
        "startSeconds": 0.0,  
        "endSeconds": 3.5  
      },  
      {  
        "startTime": "00:13:05.9900000",  
        "endTime": "00:13:14",  
        "startSeconds": 786.0,  
        "endSeconds": 794.0  
      }  
    ],  
    "seenDurationRatio": 0.014496221662468516  
  }  
]
```

Video Indexer extracted metadata snippet

“annotations”:

```
[{"name": "man", "timeRanges": [{"start": "00:00:00.0660000", "end": "00:00:17.1330000"}]}, {"name": "person", "timeRanges": [{"start": "00:00:00.0660000", "end": "00:00:17.1330000"}, {"start": "00:00:18.1990000", "end": "00:00:27.1500000"}]}, {"name": "red", "timeRanges": [{"start": "00:00:00.0660000", "end": "00:00:05.4000000"}, {"start": "00:00:08.5990000", "end": "00:00:10.7330000"}, {"start": "00:00:11.7990000", "end": "00:00:15"}]}, {"name": "wearing", "timeRanges": [{"start": "00:00:01.1330000", "end": "00:00:02.2000000"}, {"start": "00:00:06.4660000", "end": "00:00:07.5330000"}, {"start": "00:00:13.9330000", "end": "00:00:15"}]}, {"name": "suit", "timeRanges": [{"start": "00:00:01.1330000", "end": "00:00:02.2000000"}, {"start": "00:00:03.2660000", "end": "00:00:08.6000000"}, {"start": "00:00:12.8660000", "end": "00:00:17.1330000"}]}, {"name": "screen", "timeRanges": [{"start": "00:00:05.3990000", "end": "00:00:06.4660000"}, {"start": "00:00:23.5330000", "end": "00:00:24.6000000"}, {"start": "00:00:26.7330000", "end": "00:00:27.1500000"}]}, {"name": "newspaper", "timeRanges": [{"start": "00:00:18.1990000", "end": "00:00:27.1500000"}]}, {"name": "sign", "timeRanges": [{"start": "00:00:20.3330000", "end": "00:00:21.4000000"}, {"start": "00:00:23.5330000", "end": "00:00:27.1500000"}]}
```

Video Indexer insights widget

```
<iframe width="580" height="780"
src="https://www.videoindexer.ai/
embed/insights/e2131c8c8b"
frameborder="0"
allowfullscreen></iframe>
```

The screenshot displays the 'Cognitive Insights' interface for a video. At the top, there are navigation options for 'Cognitive Insights', 'Transcript', and 'Search', along with a language dropdown set to 'English'. Below this, a 'People' section shows a row of ten circular profile icons. The first icon is highlighted, leading to a detailed view for 'Scott Guthrie'. This view includes a profile picture, a 'Show biography' button, and a 'Find in Bing' button. A progress bar below the name indicates that Scott Guthrie appears for a total of 23% of the entire duration. The 'Keywords' section features a collection of buttons for terms such as 'Hybrid', 'platform', 'Protection', 'document', 'company', 'transformation', 'experience', 'Windows Defender', 'adobe', 'attack', 'Team Services', 'Visual Studio family', 'Developer services for teams', 'Fortune', and 'Microsoft Intune'. At the bottom, the 'Speech Sentiment' section shows a legend with three categories: 'Neutral (29%)', 'Positive (70%)', and 'Negative (1%)', accompanied by a horizontal bar chart where the majority of the segments are green, representing positive sentiment.

Video Indexer roadmap

- New Video AI algorithms
 - Face emotions, Content moderator, audio event detection, more objects in object model, activity detection.
- Customization of Models
 - Enable custom models for targeted output
 - Includes - Integration of CRIS, evolution of face recognition, IRIS for videos.
- Live video support
 - Integrate with live channels to provide near real time metadata from the video stream
 - Auto Transcription, Live Translation, Real-time face detection
- Expanded Integration
 - Continued improvements to our Logic Apps/Flow Connector
 - Support for Triggers
 - Simplified integration with Cosmos DB and Azure Search

Azure Media Services, Azure Logic App & Azure Functions

"Serverless Automation"

Example Logic App

- Upload Video File to Blob Storage from OneDrive
- Encode, Index & Summarize
- Publish and Share streaming URL via SMS, Email, and Twitter

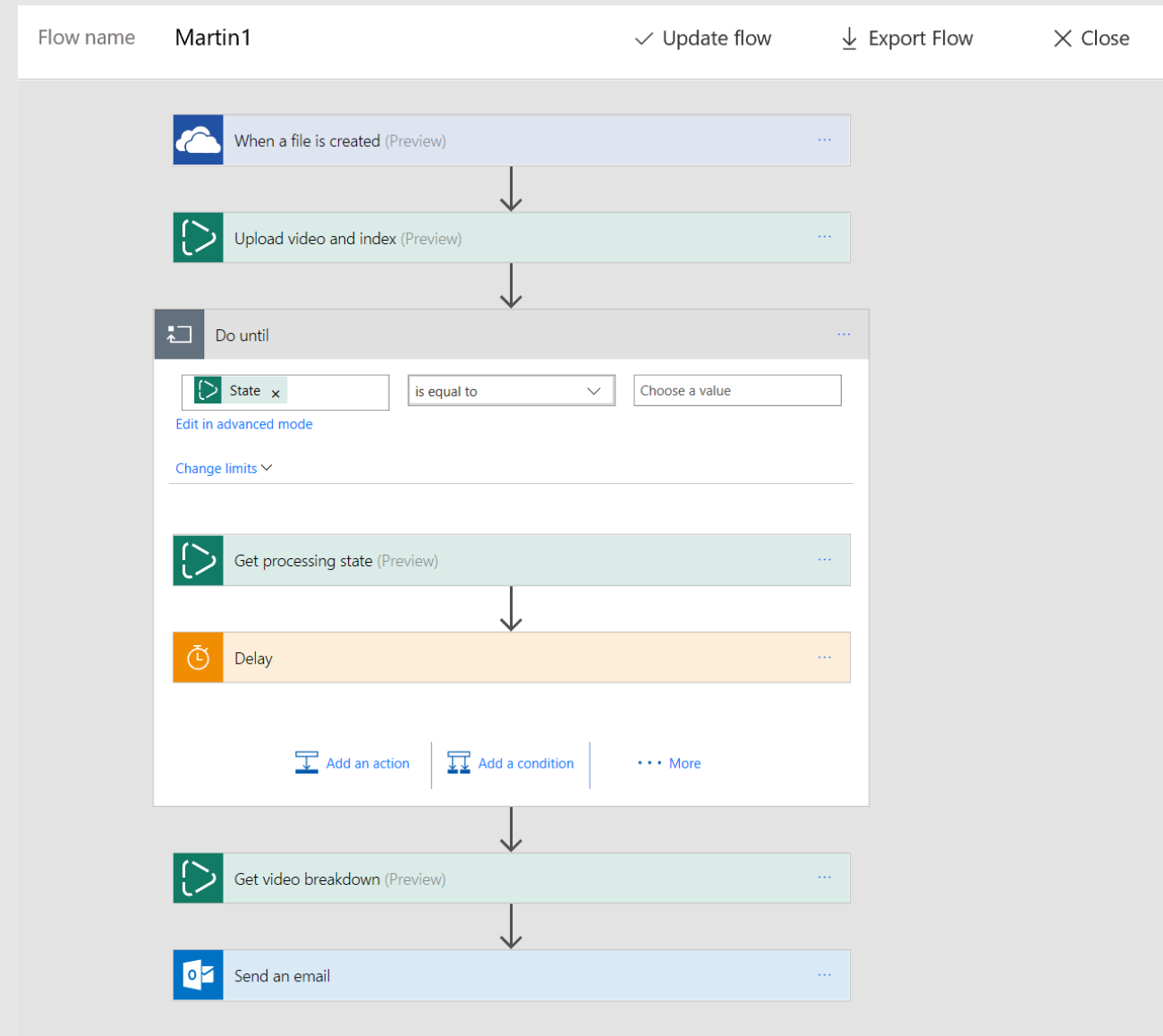


Azure Flow

“Quick development”

Example Flow App

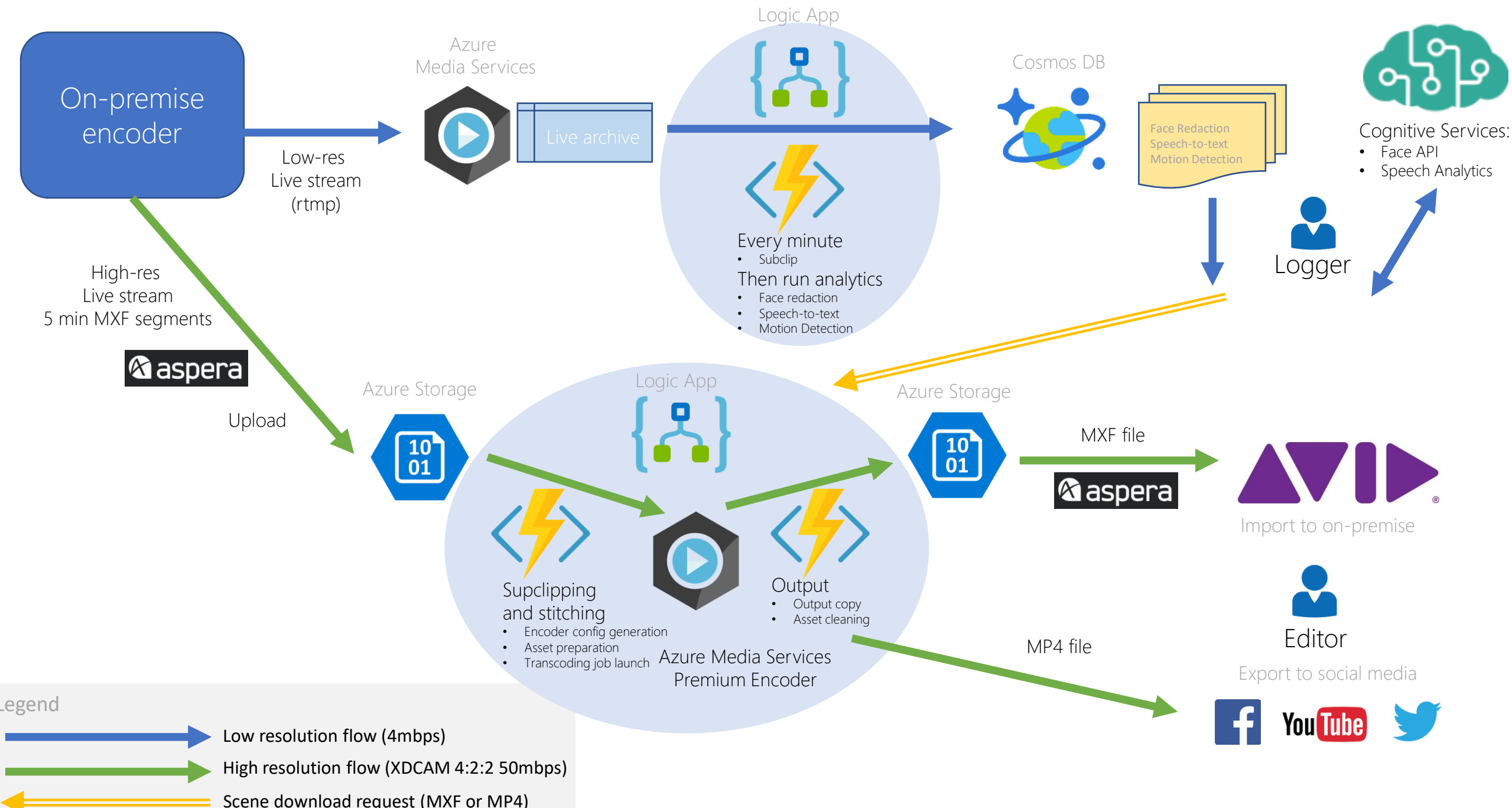
- Upload Video File to Blob Storage from OneDrive
- Send to Video Indexer
- Publish and Share streaming URL via SMS, Email, and Twitter



Getting creative with combining....

Logic Apps/Functions +
Live Streaming +
Live Subclipping +
Media Analytics +
AMP Plugin
+ Bing Translation

**Auto Face &
Object Tagging/
Live Captioning/**
+
Live Translation



On-premise encoder

Low-res Live stream (rtmp)

High-res Live stream 5 min MXF segments



Upload



Azure Storage



Logic App

Supclipping and stitching

- Encoder config generation
- Asset preparation
- Transcoding job launch

Azure Media Services Premium Encoder

Output

- Output copy
- Asset cleaning



Azure Storage

MXF file



Import to on-premise



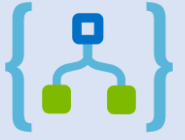
Editor

Export to social media



MP4 file

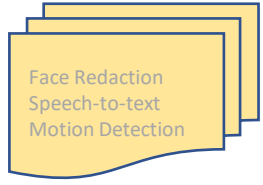
Logic App



Every minute

- Subclip
- Then run analytics
- Face redaction
- Speech-to-text
- Motion Detection

Cosmos DB



Face Redaction
Speech-to-text
Motion Detection



Cognitive Services:

- Face API
- Speech Analytics



Logger

Legend

- Low resolution flow (4mbps)
- High resolution flow (XDCAM 4:2:2 50mbps)
- Scene download request (MXF or MP4)

