

Video AI: What's in your content?

Martin Wahl
Principal Program Manager
Microsoft Azure Media Services
Redmond, WA
mawahl@microsoft.com



Azure is the most trusted and certified Cloud ...

Standards
and Guidelines



digital production partnership

Production
and Broadcast



Content Protection
Standard v3.0



Motion Picture Association

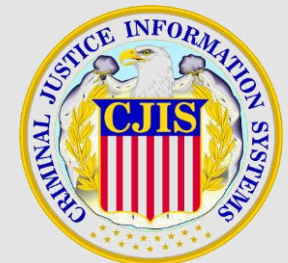
Cloud & Application
Security Guidelines



Federation Against
Copyright Theft

+ Relevant
Frameworks

...for media and entertainment workflows



Azure Media Services

Azure Media Services is an extensible, multi-tenant platform that enables end-to-end video workflows in the Azure public cloud and relies on **Azure Storage, Azure Compute, Azure Networking**, and other Azure Cloud Infrastructure Services.





Video Artificial Intelligence

Imagine if you could...

- Automatically create all of the metadata (spoken text, written text, faces, places, objects etc) for any video, including from their archives – flag any adult or objectionable content
- Relate 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 viewers to take advantage of deep search capabilities – search by Name, Search by Keyword, Search by Topic – and view content in any language even if different from source
- Distribute playable clips of live video content internally or externally without long download or encode times
- Understand which part of your broadcasts were most popular/interesting to viewer and then add Interactive Elements into the viewing experience
- Create automated summaries or highlight reels of video content based on scene detection, specific people, and motion/actions within the video

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

Video AI features



Speech-to-Text

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/Image 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 summarization

Create summaries of long videos to enable quick previews

Content moderation

Detect and prevent explicitly visual or objectionable content

Object detection

Detect objects based on a pre-defined object model

Available as individual Media Processors via APIs

Video Indexer | Bundling 20 AI Features Together



Linguistic Transcript

Convert speech to text for 10 languages



Noise canceling

Eliminate background noise for call recordings using Skype filters



Face detection

Find when does each face appears in the video



Keywords Extraction

Find out the keywords discussed in each segment



Contextual Search

Understand the context of search results



Spoken Language ID

Detect spoken language to support multi language content



Call recording enhancement

Optimized ingestion for calls



Face grouping

Identify multiple appearances of the same person



Sentiment Analysis

Compare levels of positive vs negative spoken or written moments over the timeline



In-place Editing

Make manual fixes for errors detected



Custom Vocabulary

Fit to Industry, market and domain specific terms



Annotations

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



Identification

See name, job and biography of celebrities and ordinary people



Content Moderation

Detect explicit visuals or text in audio or overlay



Sub-Clipping

Source video is stored once for multiple playlists of video segments



Speaker Diarization

Map and understand who spoke when



OCR

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



Shot Detection

Detect when a shot starts/ends based on visual analysis



Recommendations

Find more videos with similar people discussing similar topics



Programmatic APIs

Index / Search API and UI widgets enable embedding in other website/apps



Translation

Translate source to 54 languages – text or voice



Brand detection

Track brand mentions in speech or on screen overheads



Emotion sensing*

Detect emotions expressed in speech, vocal signals and facial expressions



Logos*

Identify visual logos that appear on screen

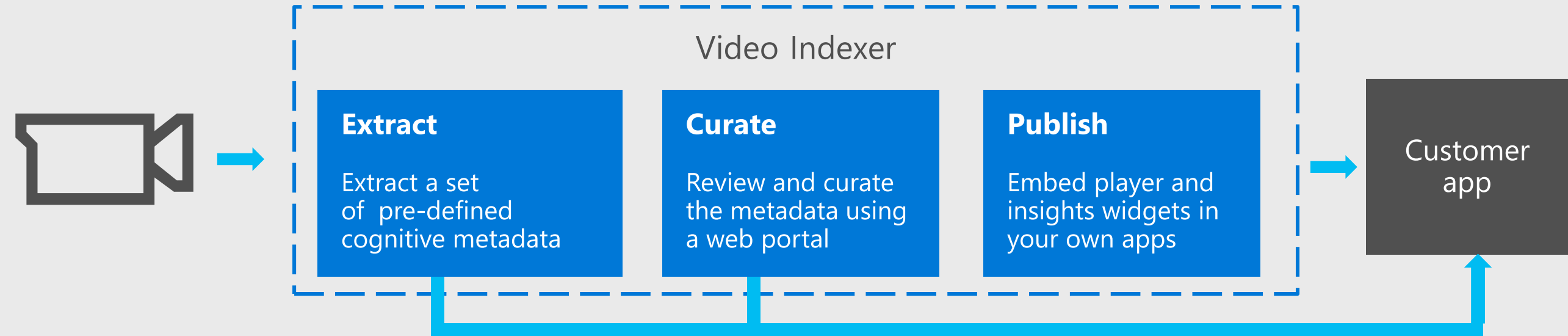


Live Analytics*

Analyze content coming from a live broadcast source

* Coming Soon

3 steps of using Video Indexer





Microsoft Video AI Demo

<http://www.videoindexer.ai>

Video Indexer | Building Blocks

Re-envisioned classic experiences:

Gallery

1 In-video Search

Player

2 Insights Panel
3 Content based Recommendations

Editor

4 Cognitive timeline
5 On the fly preview
6 Segments Playlist

APIs & Widgets

- DELETE** Delete Breakdown
- GET** Get Processing State
- GET** Get Breakdown
- GET** Get Vtt Url
- GET** Get Insights Widget Url
- GET** Search
- GET** Get Player Widget Url
- POST** Upload Video

<https://aka.ms/viapi>

AVI Platform

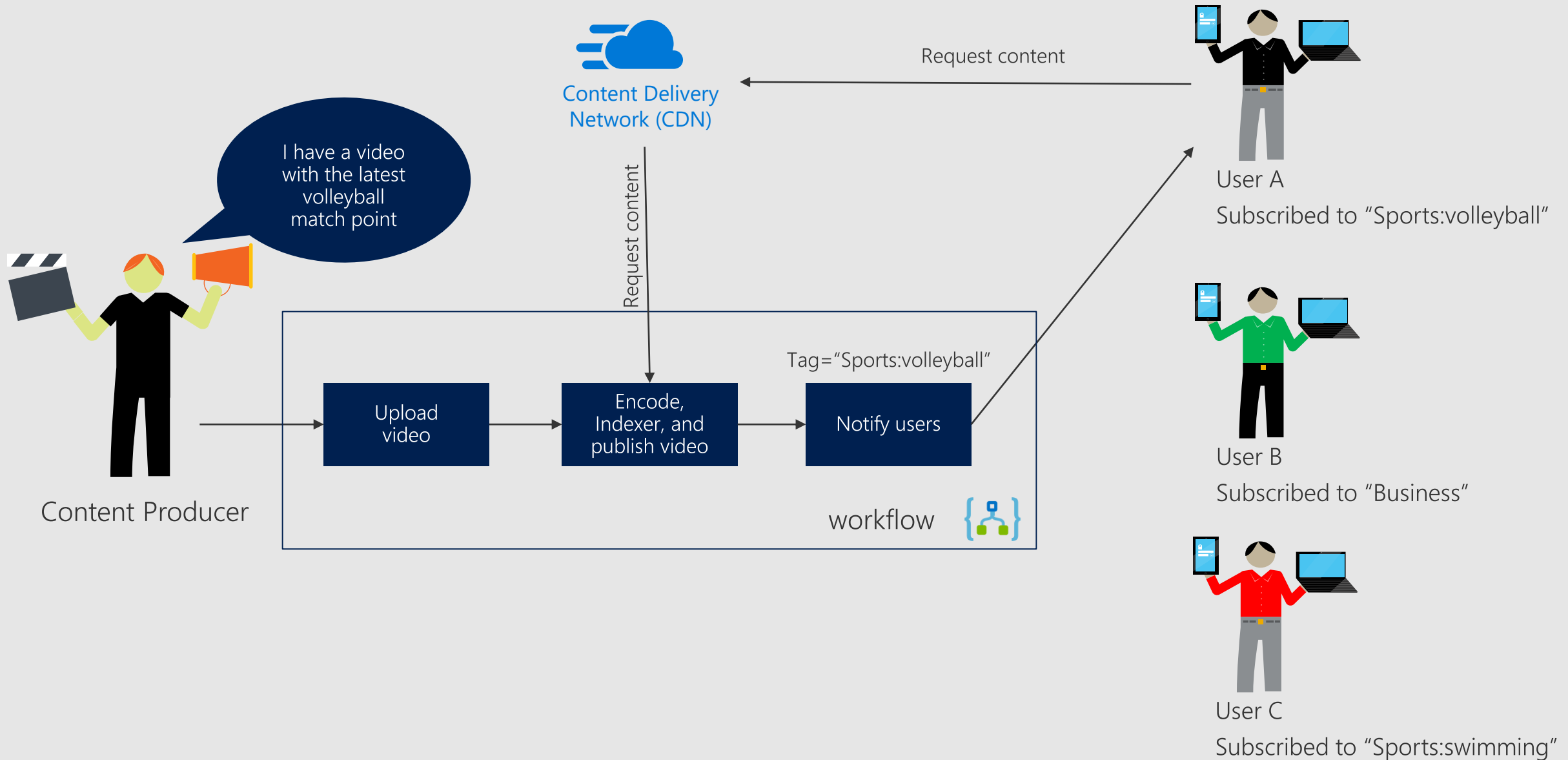


Indexing

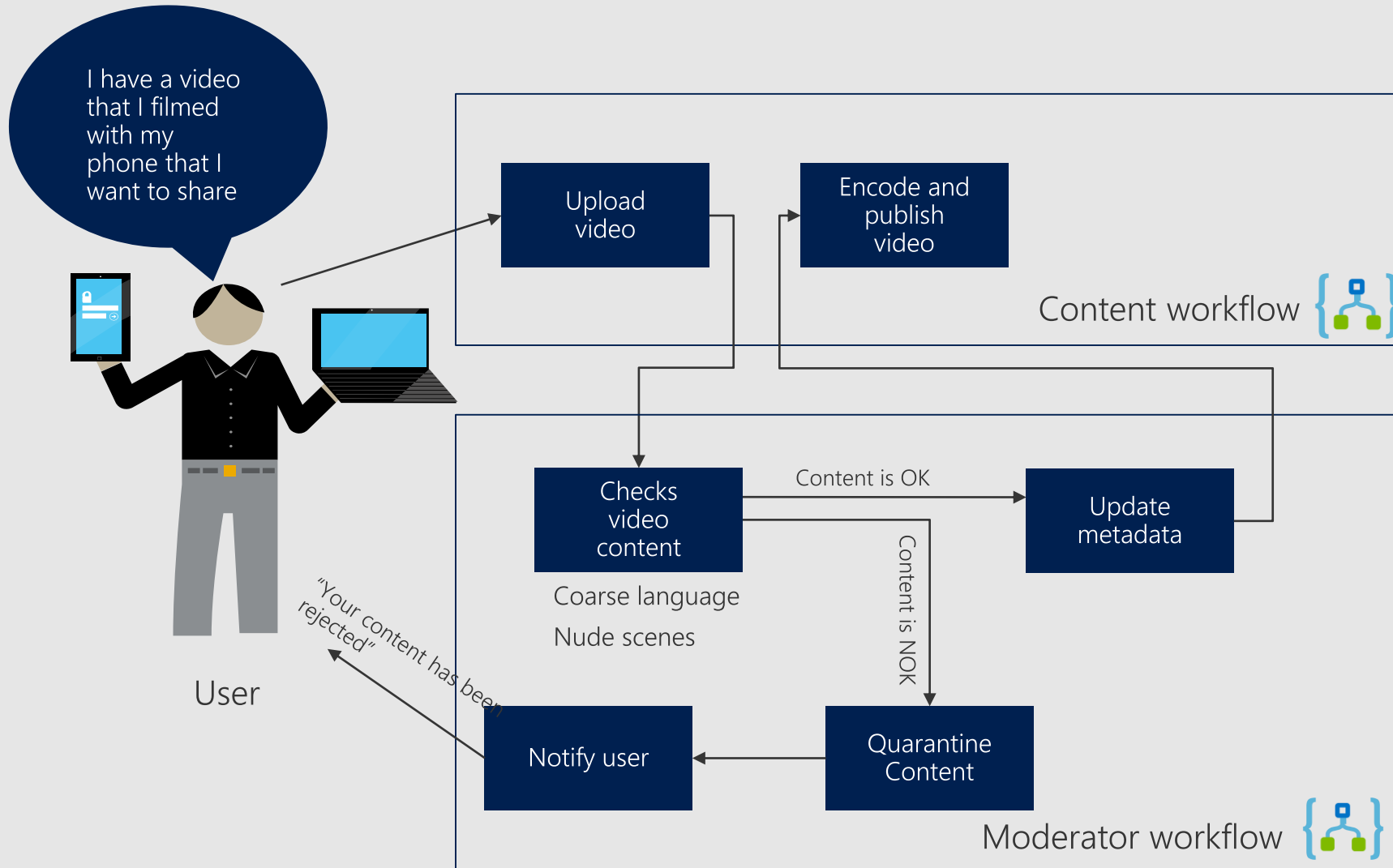
Search

Streaming

Example: Customized News Services



Example: User Generated Content Filtering



Example: Deliver Content to multiple partners

I have a list of partners who need to receive specific content files in different formats and qualities.



I also have a list of assets which need to be distributed to those partners.



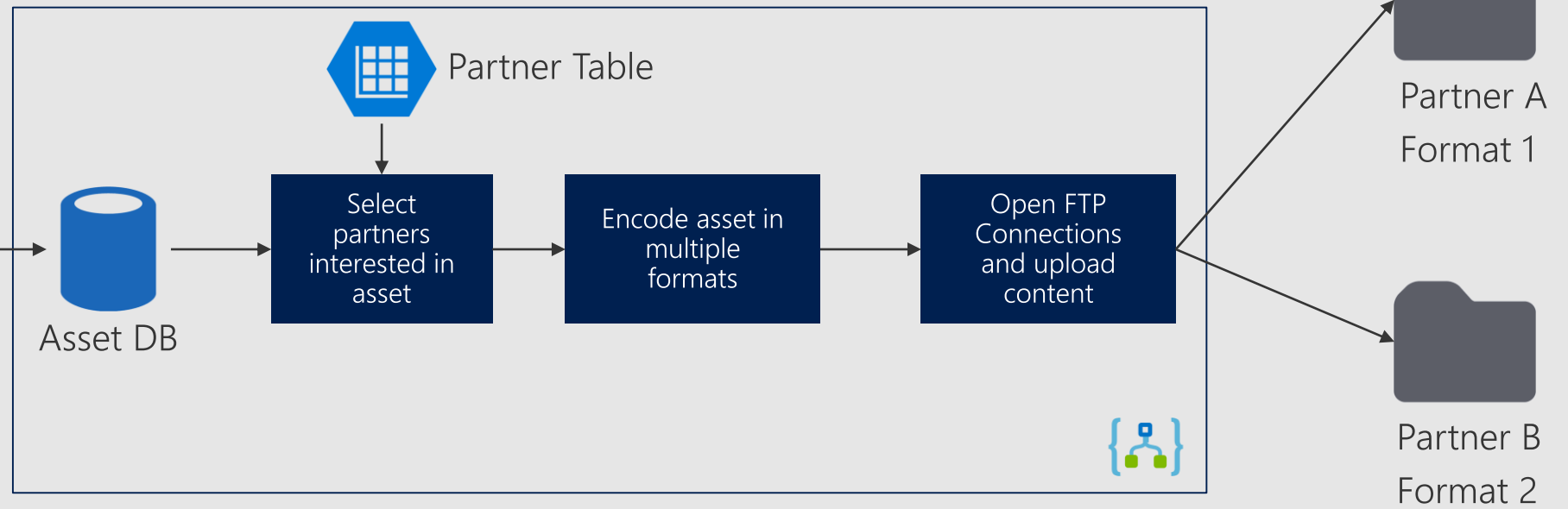
If I add/remove an asset, the system needs to distribute/remove it to/from the right partners in the corresponding formats.



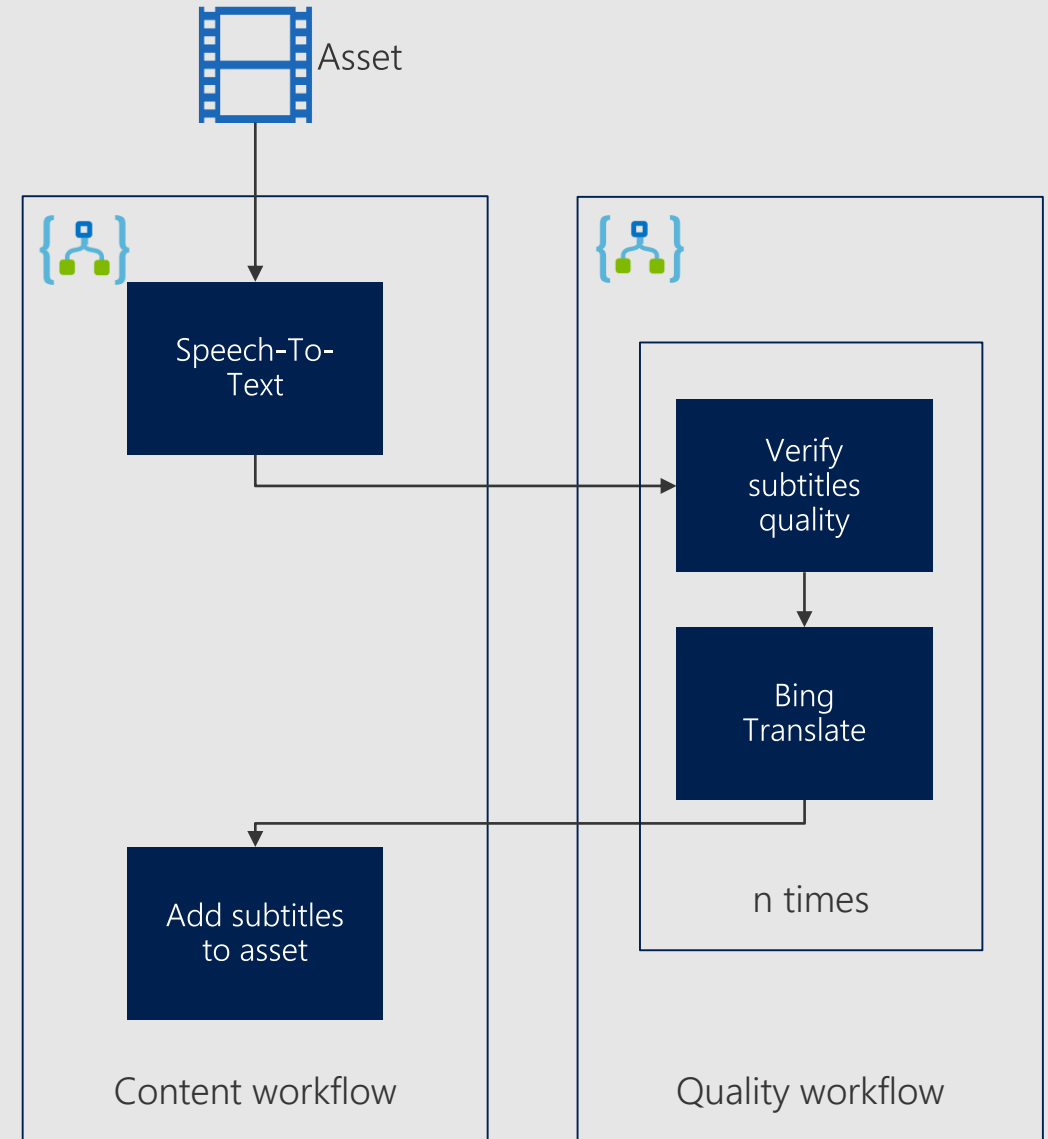
If I add/remove a partner, the system must distribute or delete the right assets to/from the destination.



Asset



Example: Auto Generate subtitles and translate



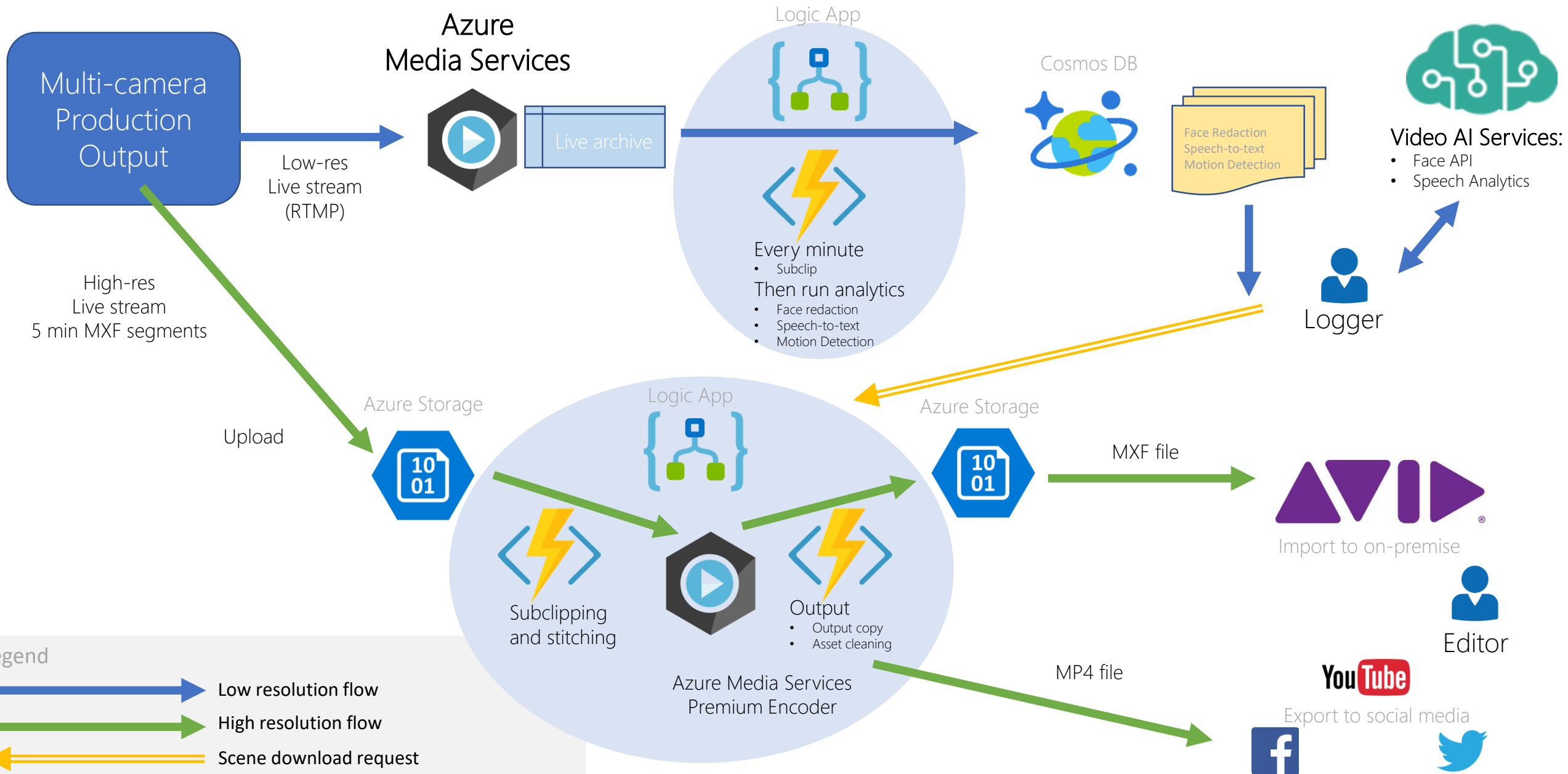
Big Brother Going Into The Cloud

“Adopting this ground-breaking technology will completely revolutionize the way we produce our global formats and opens up an unprecedented level of creative freedom. This is the first of many innovations we are working on that will optimize our productions around the world, giving our audiences across multiple platforms a much richer experience. [This] isn't about cost cutting or optimization, this is actually about **being able to tell great stories in totally different ways.**”

— Lisa Perrin
CEO, Creative Networks, EndemolShine



Using Cloud & Video AI within a Production Environment





Martin Wahl
Principal Program Manager
Microsoft Azure Media Services
Redmond, WA
mawahl@microsoft.com