

Using Advanced Video AI in Media to answer...

“What’s In My Content?”

Martin Wahl
Principal Program Manager, Cloud AI & Cognitive Services
Redmond, WA
mawahl@microsoft.com





The AI transformation

Over \$150B in IT Spend expected (\$19B in Media & Entertainment)



Digital Agents

Transform the engagement with your customers
(Prediction, Customization, Support, Upsell)



Intelligent Apps

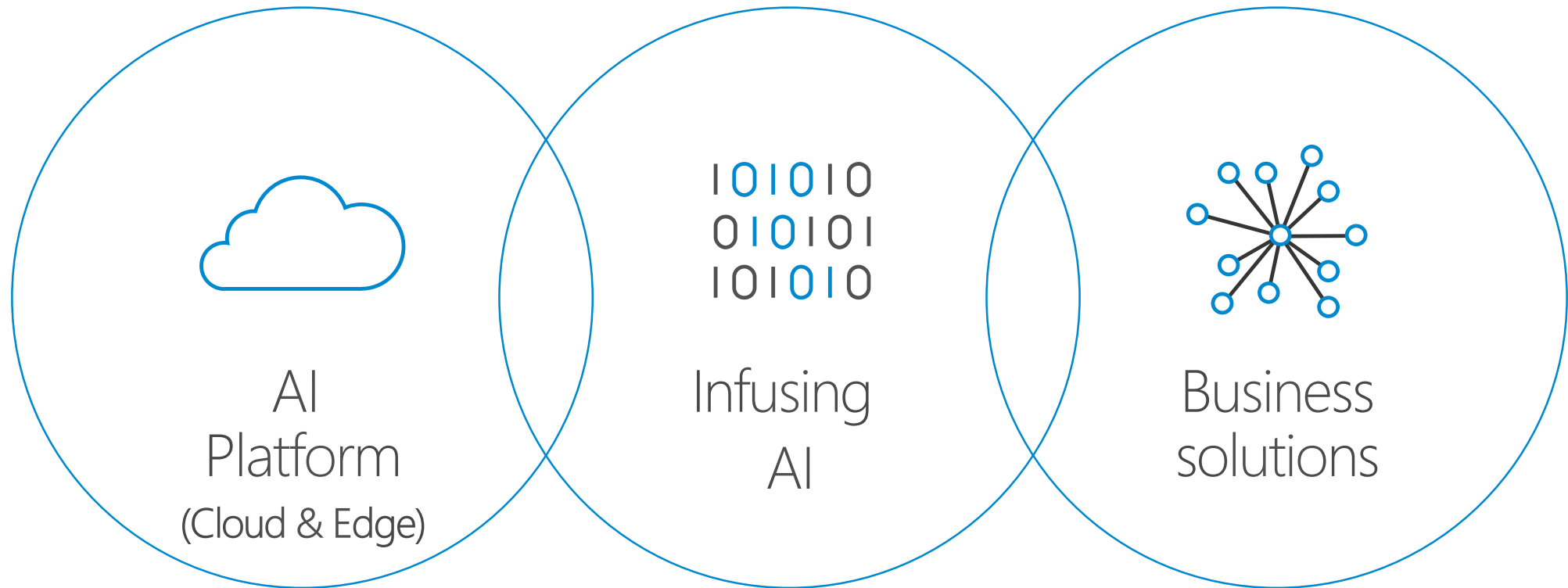
Leverage AI to create the future business applications



AI-driven Business processes

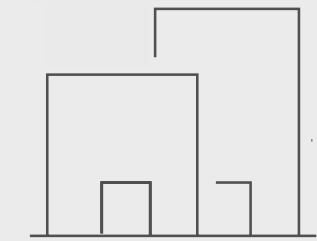
Transform critical business processes with AI and empower your employees

Microsoft AI Investment Areas



AI at the Edge; AI in the Cloud

AI over any data, anywhere



On-premises data

← Hybrid →



Cloud data

1

Productive
AI platform in the
Cloud & at the Edge

2

Comprehensive
Enterprise data estate

3

Enterprise-proven
solutions



LOB



CRM



Graph



Image



Social



Device /
IoT

Microsoft AI Platform

Azure AI Services

PRE-BUILT AI

Cognitive Services

CONVERSATIONAL AI

Bot Service



CUSTOM AI

Azure Machine Learning

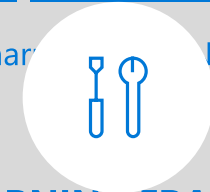
CODING & MANAGEMENT TOOLS

VS Tools for AI

Azure ML Studio

Azure ML Workbench

Others (PyCharm, Jupyter Notebooks...)



DEEP LEARNING FRAMEWORKS

3rd Party

Cognitive Toolkit

TensorFlow

Caffe

Others (Scikit-learn, MXNet, Keras, Chainer, Gluon...)

Azure Infrastructure

AI ON DATA

AI COMPUTE

Cosmos DB

SQL DB

SQL DW

Data Lake

HDInsight

DSVM

Batch AI

ACS

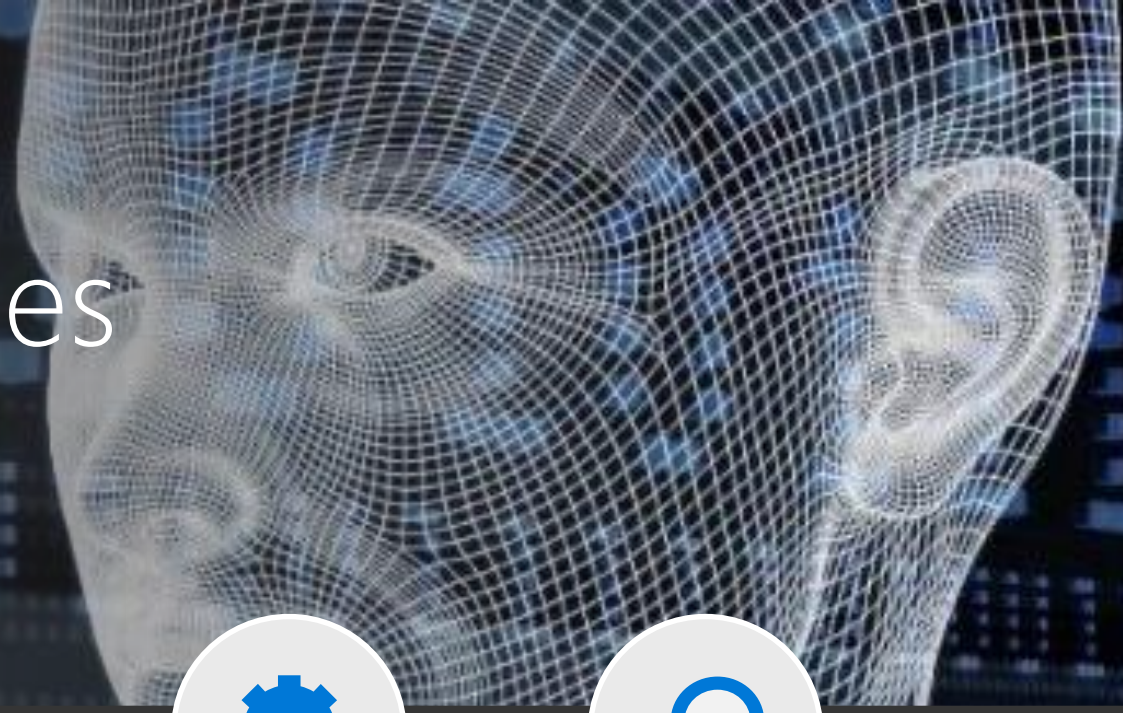
IoT Edge



CPU, FPGA, GPU

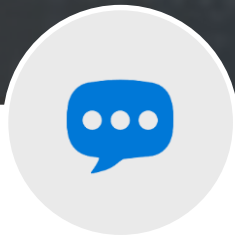
Microsoft Cognitive Services

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



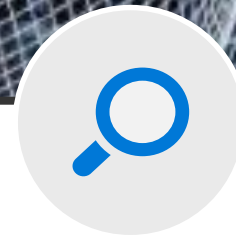
Language

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

Microsoft Cognitive Services

Give your apps a human side



Vision

Computer Vision
Content Moderator
Emotion
Face
Video Indexer



Speech

Bing Speech
Speaker
Recognition



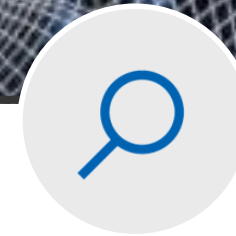
Language

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



Knowledge

Academic Knowledge
Entity Linking
Knowledge
Exploration
Recommendations
QnA Maker



Search

Bing Autosuggest
Bing Image Search
Bing News Search
Bing Video Search
Bing Web Search
Bing Entity Search

CUSTOMIZATION

Custom Vision
Service

Custom Speech
Service

Language
Understanding

Custom Decision
Service

Bing Custom
Search

Imagine if you could...

- Automatically transcribe your video or audio content
- Flag any adult or objectionable content before you publish to web sites
- Enable true video search – for any spoken word, face, object, topic, or even a product – across your entire video archive
- Create captions for your video content – in any language – and then distribute your videos to anyone, anywhere in the world, and playable on any device
- Understand which part of your video were most popular/interesting to your viewers
- Create automated summaries or highlight reels of video content based on specific people, topics, or scenes 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
- Augment Manual Labor
 - Auto Transcripts
 - Find Footage Quickly
 - Automatically Create Highlight Reels



Video AI Services

Speech-to-Text & Translation

Convert audio to text based on acoustic language models and translate to any language

Face detection & recognition

Find when each face appears in the video * identify from well known sources or custom models

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

Emotion & Sentiment Analysis

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

Video summarization

Create summaries of long videos to enable quick previews based on key metadata

Content moderation

Detect and prevent explicitly visual or objectionable content

Object detection & Recognition

Detect objects based on a pre-defined object model

Video Indexer

<http://video.ai>





Microsoft CEO Satya Nadella on company's future

Edit

Public Created 12 hours ago by Martin Wahl Report



More videos with similar people and keywords



Microsoft CEO Satya Nadella on company's future



Satya Nadella's Build 2018 keynote Shifting to intelligent edge



Royal wedding What does Meghan Markle's addition to the royal family mean for the monarchy?



Media and AI - Sponsored by NetApp, Microsoft and the Harvard Entrepreneurs Association



Seattle's Space Needle gets \$100M+ makeover Created 11 hours ago

Insights Transcript

Search... English

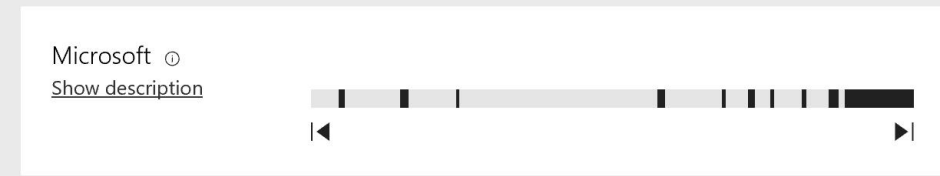
Labels

Show less

- person indoor man wall standing
- floor suit sitting outdoor woman
- table television wearing people
- electronics grass screen ground
- glasses group cellphone sky
- screenshot posing building monitor
- looking phone ceiling tree living room
- oranges water preparing cabinet music shirt
- window holding

Brands

- Microsoft
- Google
- iPhone



Speech sentiment



Edit Off | Autoscroll On | OCR On

manager Angela Mills who is legally blind showed me microsoft's new seeing

AI app so if I know hold it up it helps her read text Karen Thompson off.

You

Speaker #4



recognize objects. Here goes 350 or large orange that is oranges yeah

right I'm going to take picture care in even identify people 49 year old man with Brown hair looking happy.

Wow it left out tall and handsome but that is very first.

That

Microsoft app is running on an Apple

OCR GOLDEN CAJUN

OCR SESAME STICKS

iPhone that sentence would have been

OCR SESAME

Edit Off | Autoscroll On

Gerente Angela Mills que es legalmente ciega me mostró Microsoft nuevo ver

AI App así que si sé sostenerlo que ayuda a leer texto Karen Thompson apagado.

Que

Speaker #4



reconocer objetos. Aquí va 350 o naranja grande que es naranjas sí

derecho que voy a tomar el cuidado de la imagen en incluso identificar a la gente de 49 años de edad hombre con el pelo castaño buscando feliz.

Wow lo dejó alto y guapo, pero eso es muy primero.

Que

Microsoft APP se ejecuta en una manzana

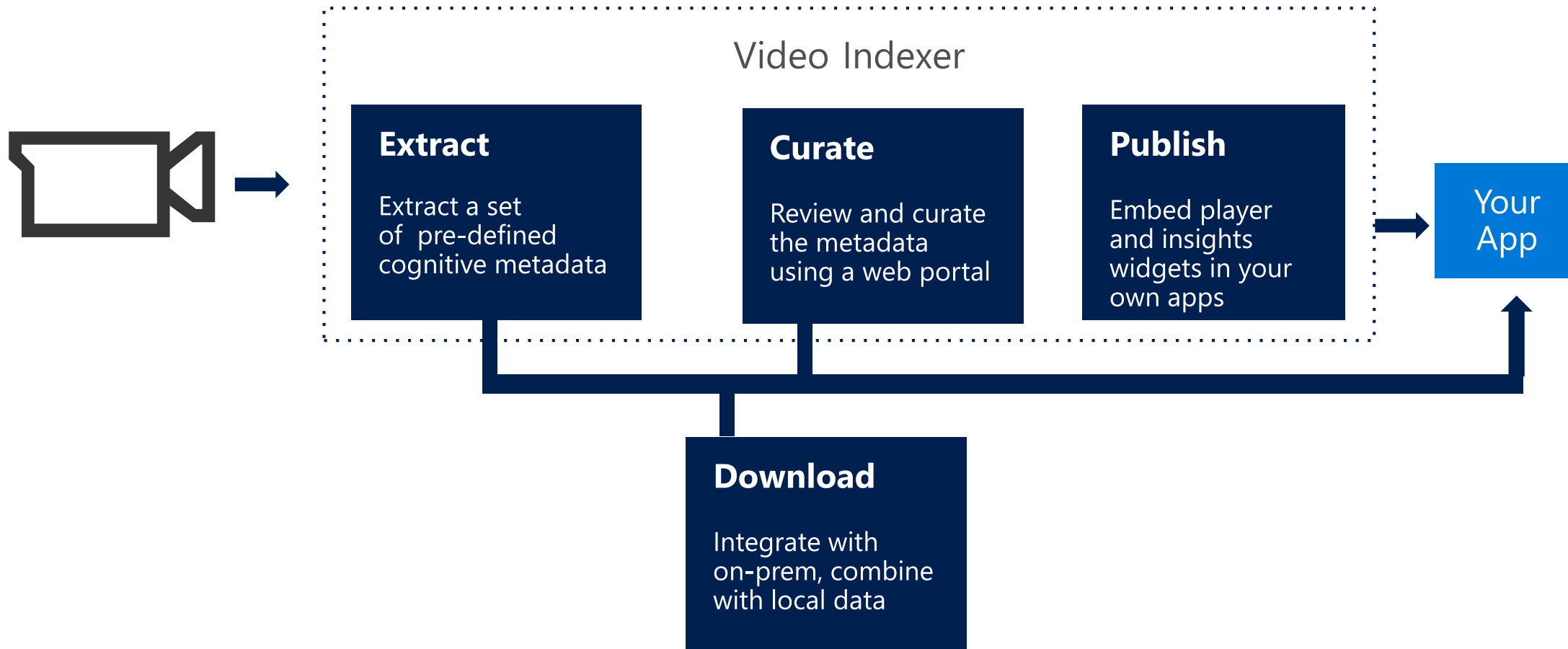
OCR Golden Cajun

OCR Palitos de sésamo

iPhone que la sentencia habría sido

- Norwegian
- Persian
- Polish
- Portuguese
- Romanian
- Russian
- Samoan
- Serbian (Cyrillic)
- Serbian (Latin)
- Slovak
- Slovenian
- Spanish**
- Swedish

Video Indexer | Typical Workflow



Video Indexer JSON insights

```
{
  "accountId": "247fc432-b3e0-4506-a849-cdee92ba3d02",
  "id": "3cd5ab1090",
  "partition": null,
  "name": "Sinclair KOMO Full newscast",
  "description": null,
  "userName": "Martin Wahl",
  "createTime": "2018-03-17T19:42:32.594073+00:00",
  "organization": "microsoft.com",
  "privacyMode": "Public",
  "state": "Processed",
  "isOwned": true,
  "isEditable": true,
  "isBase": true,
  "durationInSeconds": 2361,
  "summarizedInsights": {
    "name": "Sinclair KOMO Full newscast",
    "shortId": "3cd5ab1090",
    "privacyMode": 2,
    "duration": {
      "time": "00:39:21",
      "seconds": 2361
    },
    "thumbnailUrl": "https://www.videoindexer.ai/api/Thumbnail/3cd5ab1090/bd074748-d226-4414-8f9f-c31",
    "faces": [
      {
        "id": 36227,
        "shortId": "3cd5ab1090",
        "bingId": null,
        "confidence": 0.8965,
        "name": "Molly Shen",
        "description": null,
        "title": null,
        "thumbnailId": "488cb502-7a1a-445b-8298-6d49fc34b5ef",
        "thumbnailFullUrl": "https://www.videoindexer.ai/api/Thumbnail/3cd5ab1090/488cb502-7a1a-4",
        "appearances": [
          {
            "startTime": "00:00:05.2460000",
            "endTime": "00:00:28.4570000",
            "startSeconds": 5.2,
            "endSeconds": 28.5
          },
          {
            "startTime": "00:04:07.7810000",
            "endTime": "00:04:14.4560000",
            "startSeconds": 247.8,
            "endSeconds": 254.5
          },
          {
            "startTime": "00:06:01.5130000",
            "endTime": "00:06:08.2860000",
            "startSeconds": 361.5,
            "endSeconds": 368.3
          }
        ]
      }
    ]
  }
}
```

"faces"

```
"topics": [
  {
    "name": "chewing gum",
    "appearances": [
      {
        "startTime": "00:16:28.0800000",
        "endTime": "00:16:30.5100000",
        "startSeconds": 988.1,
        "endSeconds": 990.5
      },
      {
        "startTime": "00:36:00.7800000",
        "endTime": "00:36:03.2900000",
        "startSeconds": 2160.8,
        "endSeconds": 2163.3
      }
    ],
    "isTranscript": true,
    "id": 7
  },
  {
    "name": "Bridge collapse",
    "appearances": [
      {
        "startTime": "00:00:03.7900000",
        "endTime": "00:00:10.1800000",
        "startSeconds": 3.8,
        "endSeconds": 10.2
      },
      {
        "startTime": "00:00:41.4700000",
        "endTime": "00:00:44.9000000",
        "startSeconds": 41.5,
        "endSeconds": 44.9
      },
      {
        "startTime": "00:01:51.4900000",
        "endTime": "00:01:57.6800000",
        "startSeconds": 111.5,
        "endSeconds": 117.7
      }
    ]
  }
]
```

"topics"

```
"sentiments": [...], // 3 items
"audioEffects": [
  {
    "audioEffectKey": "Speech",
    "appearances": [
      {
        "startTime": "00:00:00.3300000",
        "endTime": "00:39:20.6200000",
        "startSeconds": 0.3,
        "endSeconds": 2360.6
      }
    ],
    "seenDurationRatio": 0.9997035154595509,
    "seenDuration": 2360.2999999999997
  },
  {
    "audioEffectKey": "HandClaps",
    "appearances": [
      {
        "startTime": "00:07:00.6470000",
        "endTime": "00:07:12.6470000",
        "startSeconds": 420.6,
        "endSeconds": 432.6
      },
      {
        "startTime": "00:10:41.2780000",
        "endTime": "00:10:43.7780000",
        "startSeconds": 641.3,
        "endSeconds": 643.8
      },
      {
        "startTime": "00:18:23.7360000",
        "endTime": "00:18:26.2360000",
        "startSeconds": 1103.7,
        "endSeconds": 1106.2
      },
      {
        "startTime": "00:18:46.7360000",
        "endTime": "00:18:48.7360000",
        "startSeconds": 1126.7,
        "endSeconds": 1128.7
      }
    ]
  }
]
```

"sentiments"

"audioEffects"

```
"annotations": [
  {
    "id": 3,
    "name": "person",
    "appearances": [...], // 85 items
  },
  {
    "id": 1,
    "name": "sky",
    "appearances": [...], // 37 items
  },
  {
    "id": 8,
    "name": "suit",
    "appearances": [...], // 27 items
  },
  {
    "id": 6,
    "name": "standing",
    "appearances": [...], // 33 items
  },
  {
    "id": 25,
    "name": "tree",
    "appearances": [
      {
        "startTime": "00:01:59.0760000",
        "endTime": "00:02:00.1390000",
        "startSeconds": 119.1,
        "endSeconds": 120.1
      },
      {
        "startTime": "00:03:44.2720000",
        "endTime": "00:03:49.5850000",
        "startSeconds": 224.3,
        "endSeconds": 229.6
      },
      {
        "startTime": "00:04:15.0870000",
        "endTime": "00:04:18.2750000",
        "startSeconds": 255.1,
        "endSeconds": 258.3
      },
      {
        "startTime": "00:04:26.7750000",
        "endTime": "00:04:32.0880000",
        "startSeconds": 266.8,
        "endSeconds": 272.1
      }
    ]
  }
]
```

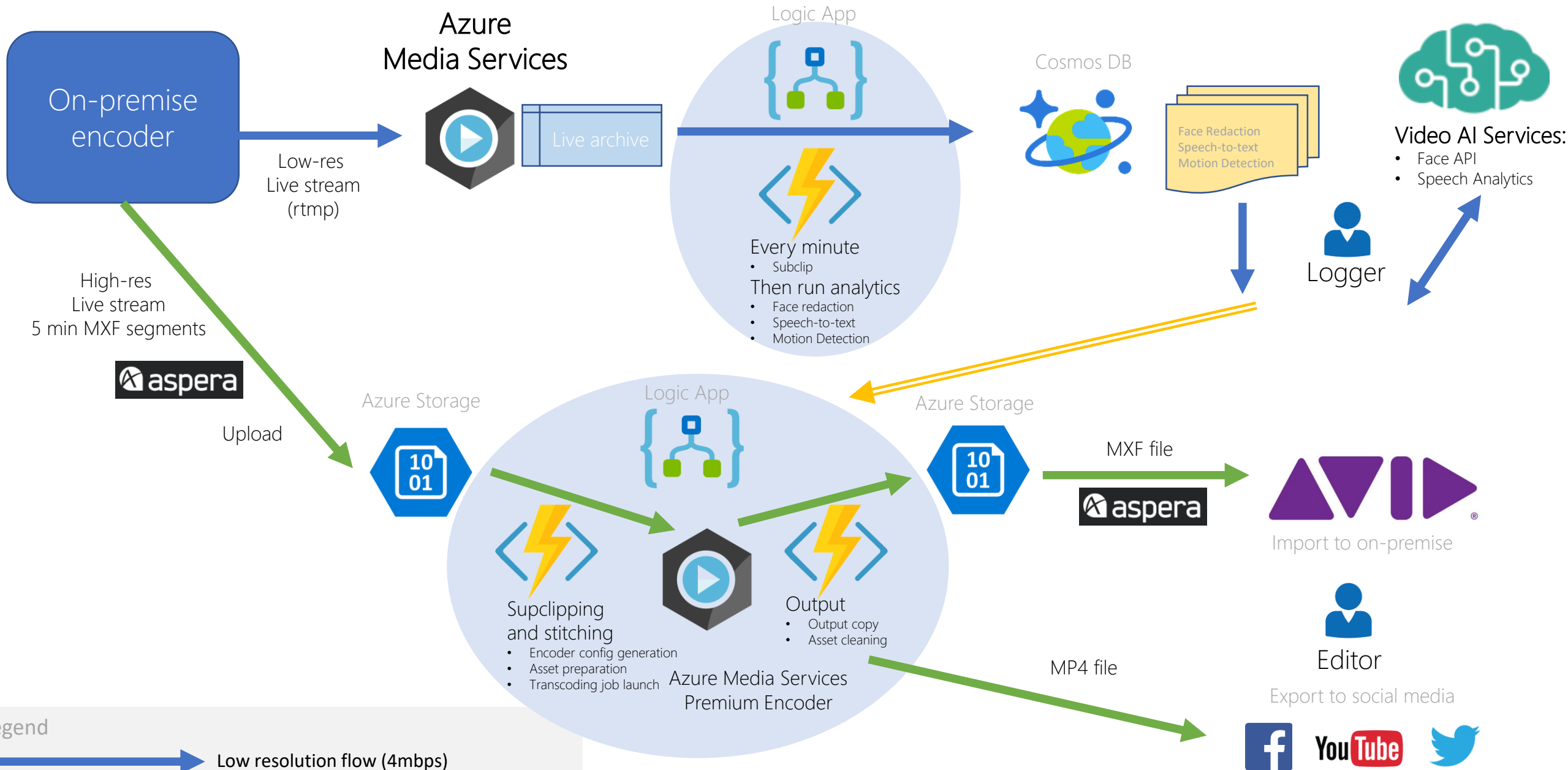
"annotations"

Big Brother Going Into The Cloud

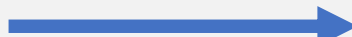


“Adopting this ground-breaking technology will completely revolutionize the way we produce our global formats and opens up an unprecedented level of creative freedom. This is the first of many innovations we are working on that will optimize our productions around the world, giving our audiences across 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

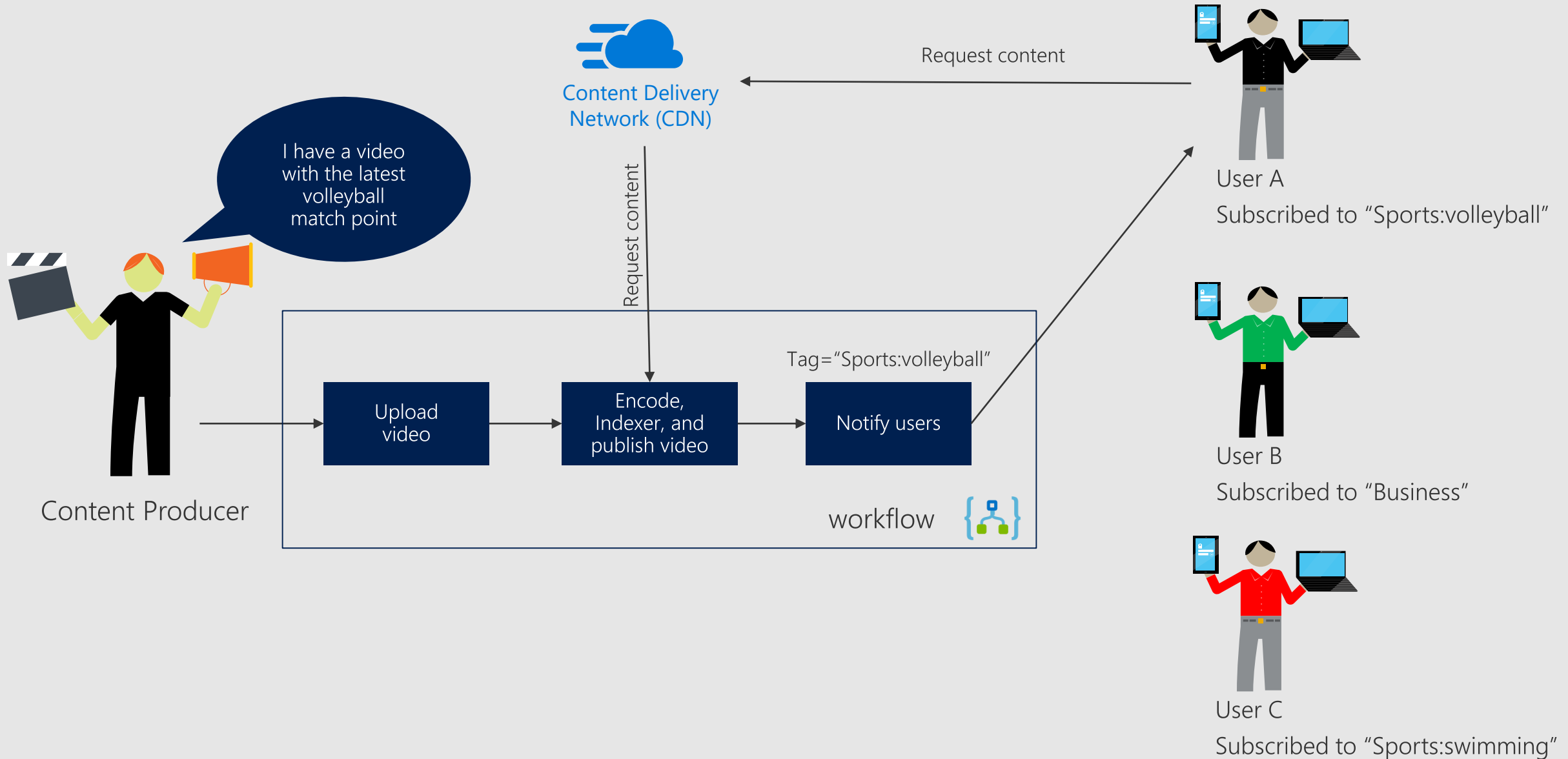




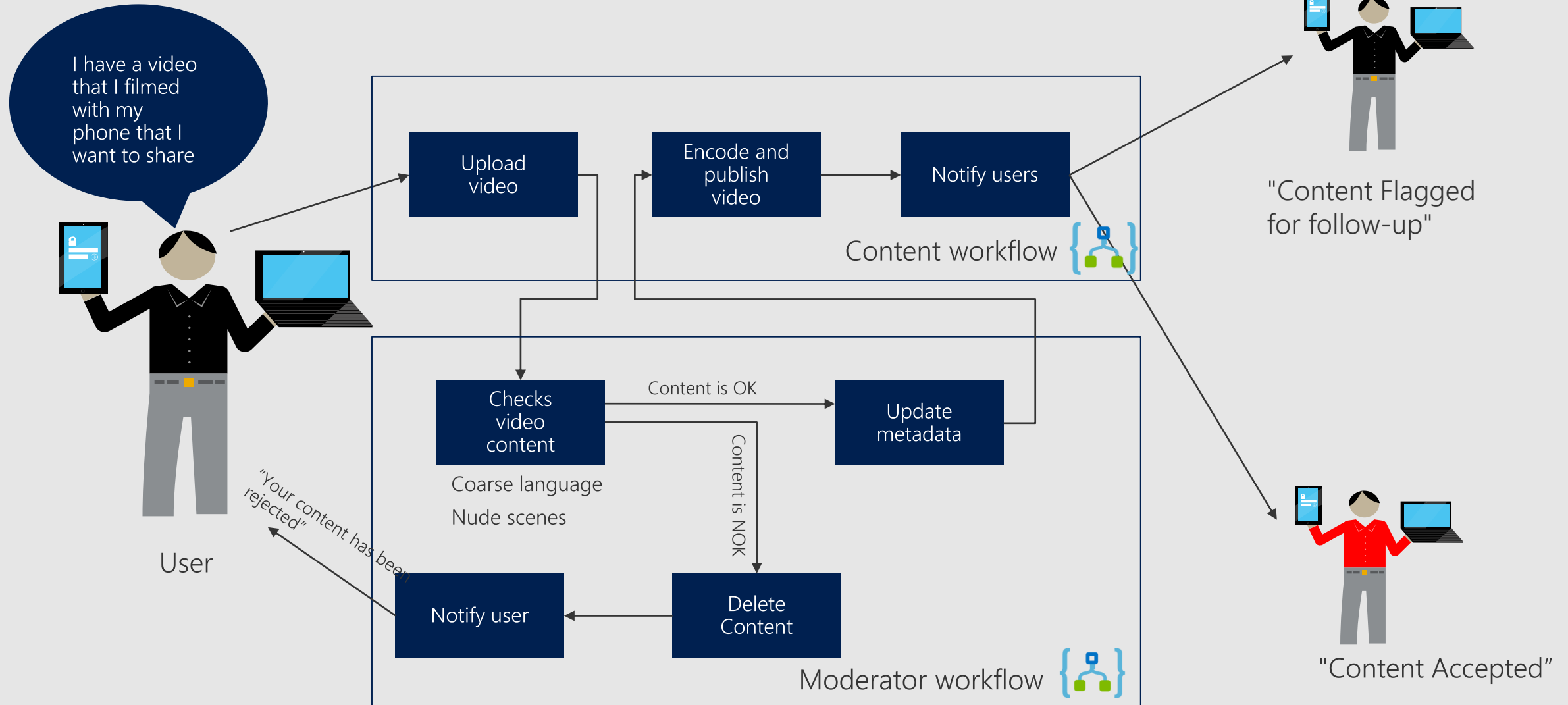
Legend

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

Example: Breaking News Notifications



Example: User Generated Content Notifications



Example 3: Deliver Content to multiple partners

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



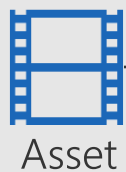
I also have a list of assets which need to be distributed to those partners.



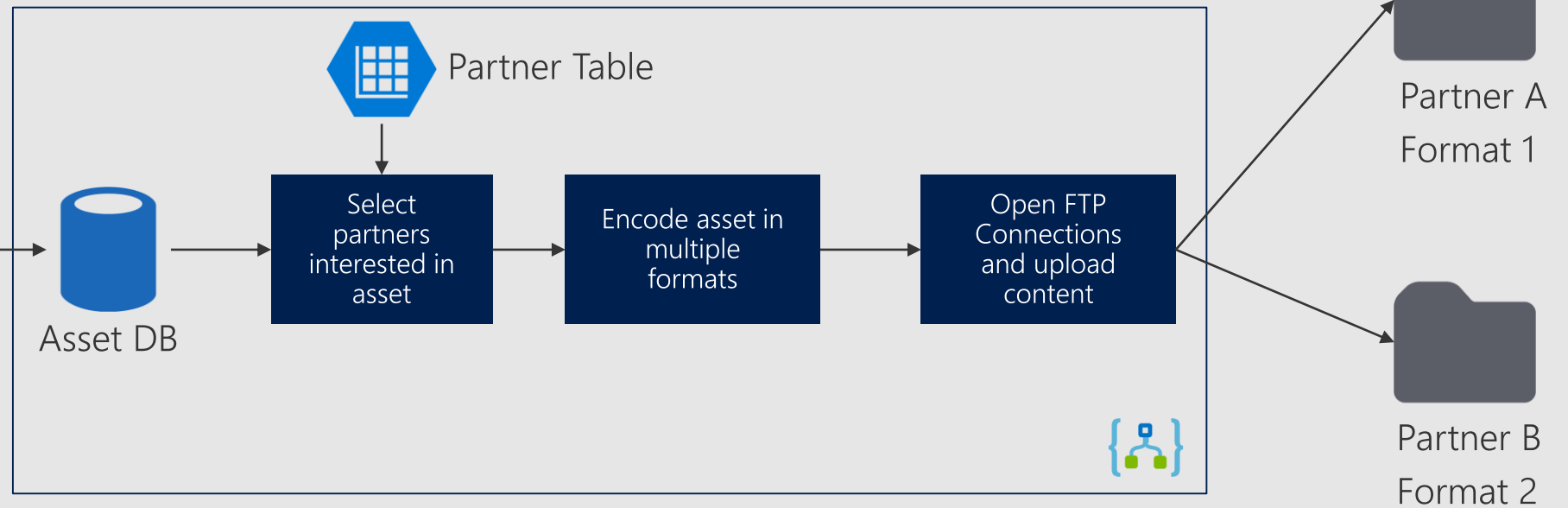
If I add/remove an asset, the system needs to distribute/remove it to/from the right partners in the corresponding formats.



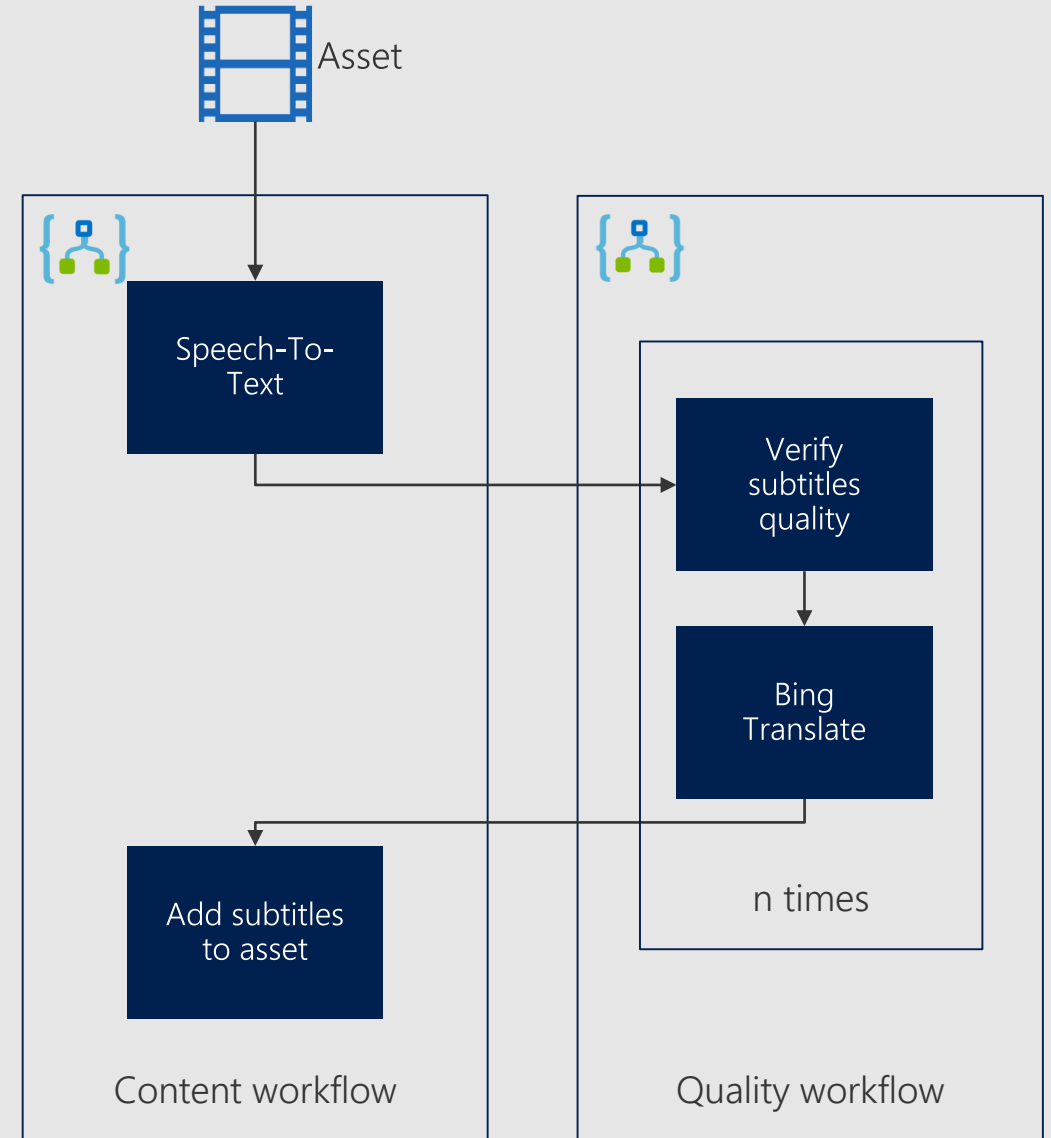
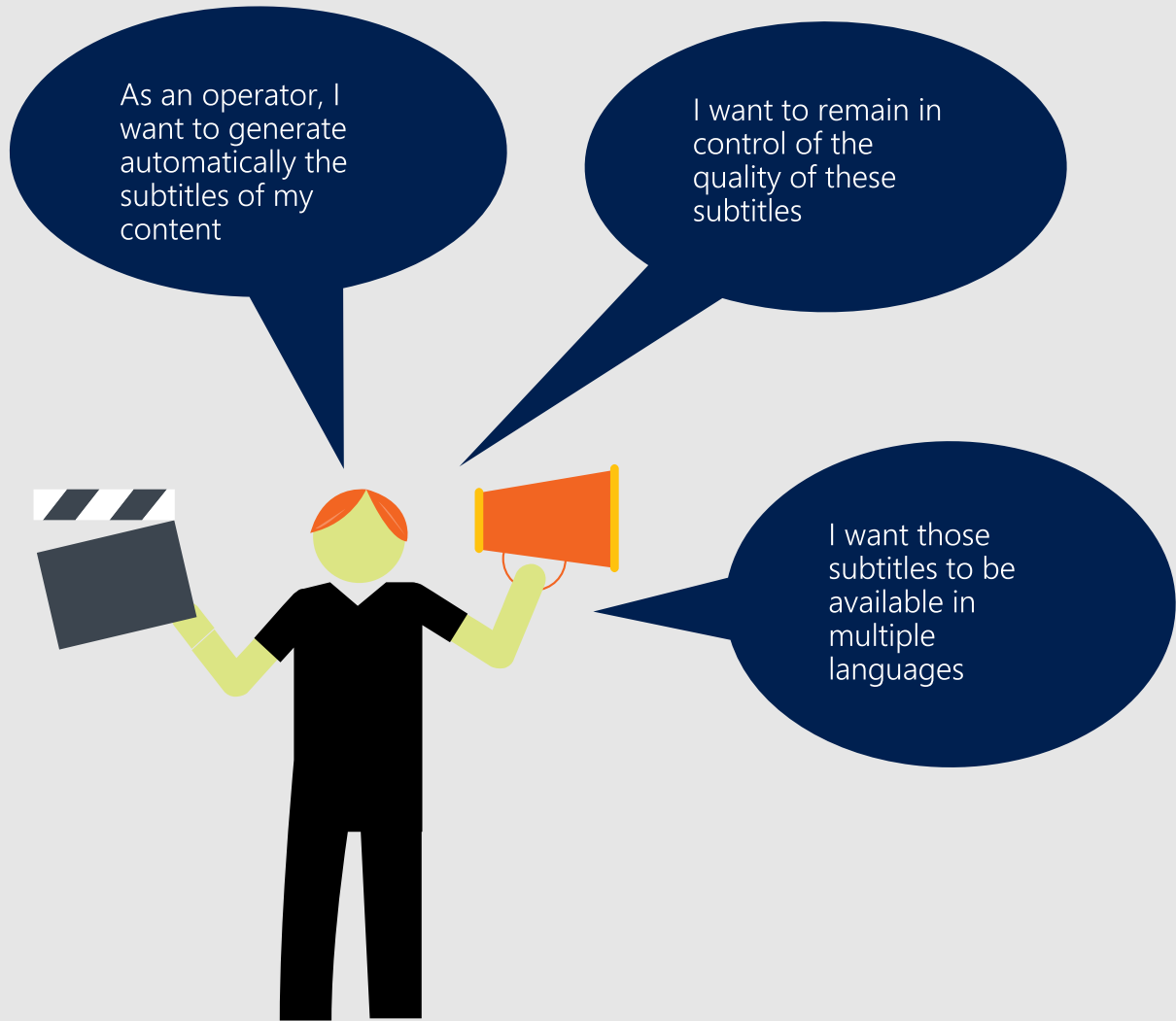
If I add/remove a partner, the system must distribute or delete the right assets to/from the destination.



Asset



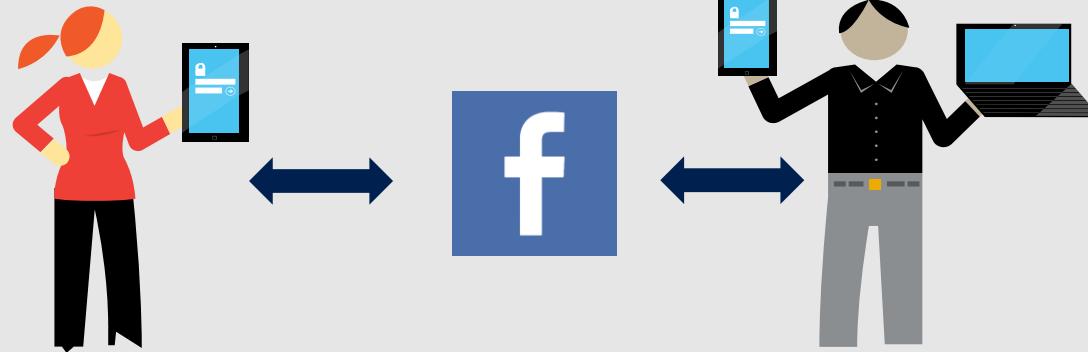
Example 4: Generating subtitles



Example 5: Recommendation engines

Here is Lisa. She is a Contoso subscriber, and she loves binging TV series.

Lisa is a Facebook friend of Fred. Fred is also a Contoso subscriber. He is a big movie fan.

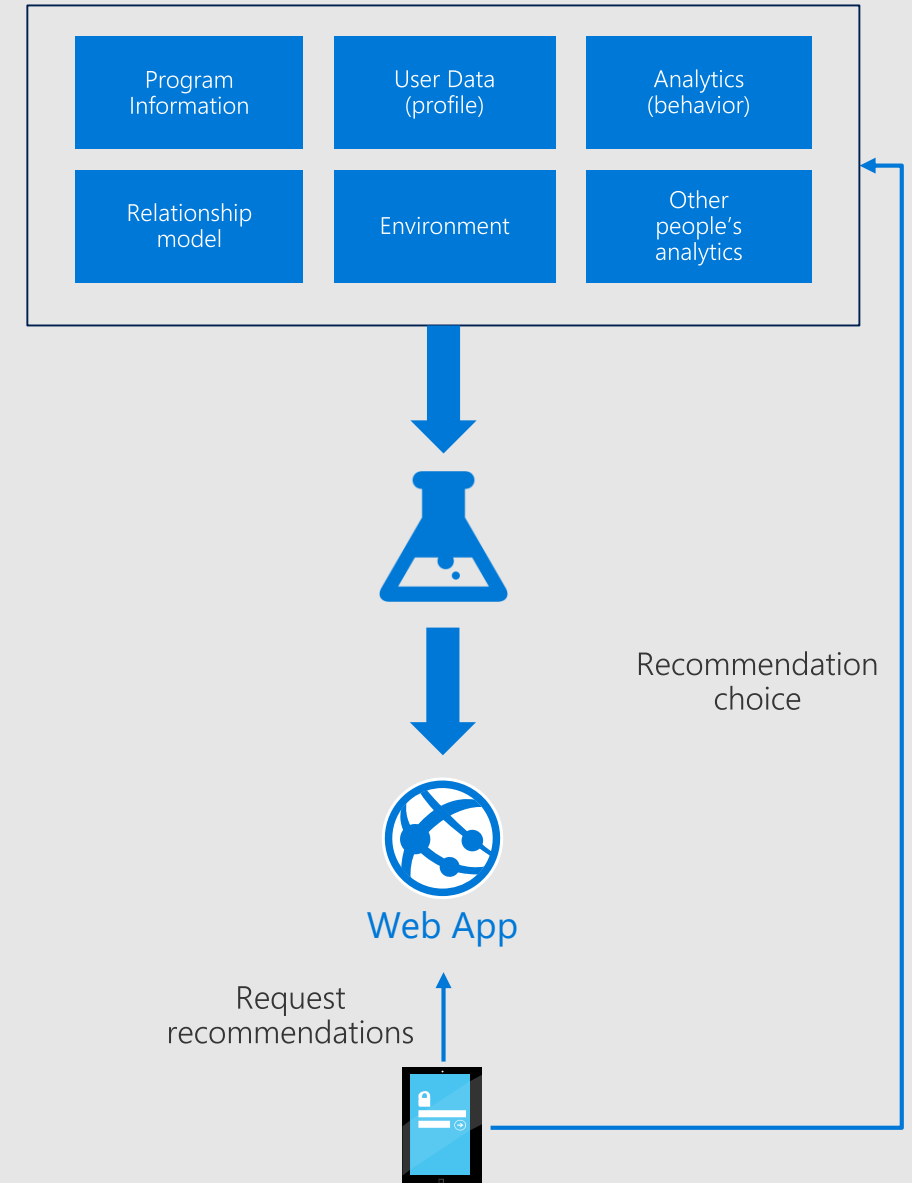


Lisa may also have some habits based on time of day, some specific channels she likes, or specific actors she loves. She may use specific devices for specific content.

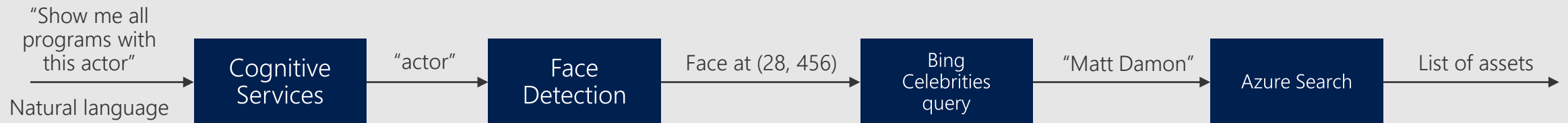
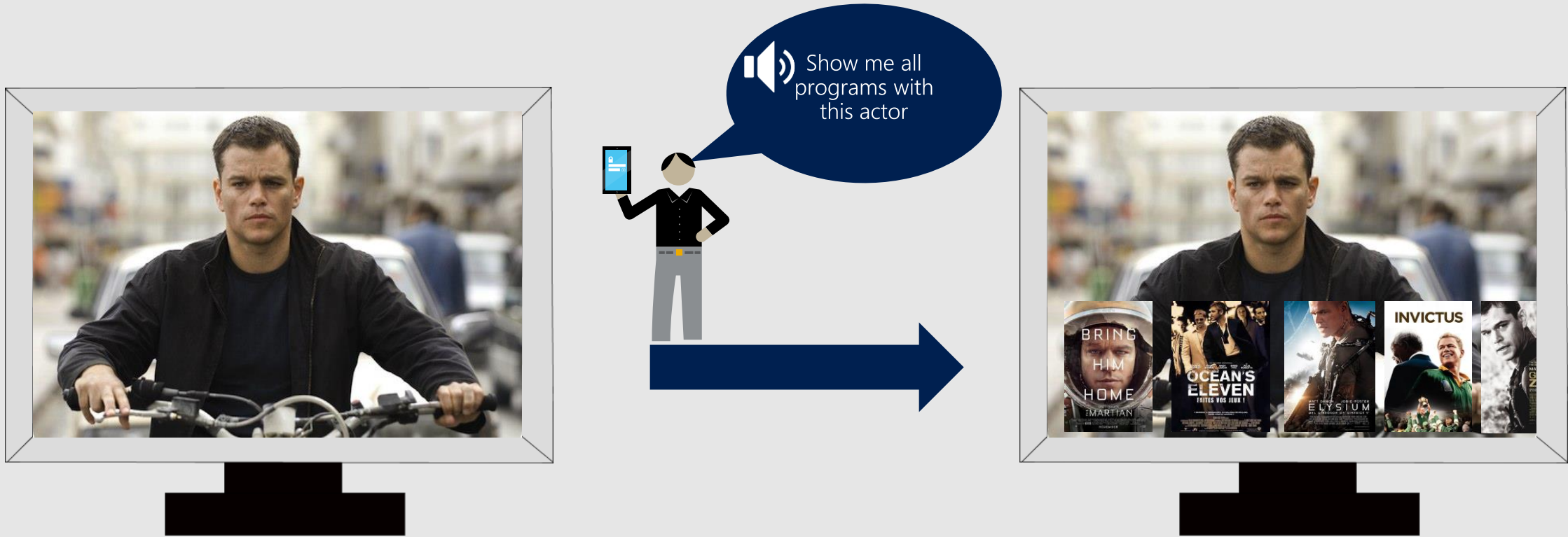
Fred and Lisa may share other common interests: actors, activities, countries, etc.

Lisa records her preferences.

Fred may be subscribed to other services, or specific interest groups.



Example 6: Transforming the user experience



Use AI/Machine Learning to answer key questions



Who are my audiences?



How many users are viewing my content now?



What devices are my customers using to consume content?



What is the peak consumption time?



What do my customers like to watch?



How do I enable my customers to find content that matches their preferences?



What content should be recommended to my users for maximum engagement?



How do I find related programs for targeted advertisements?



What is the likelihood of success of a new program?

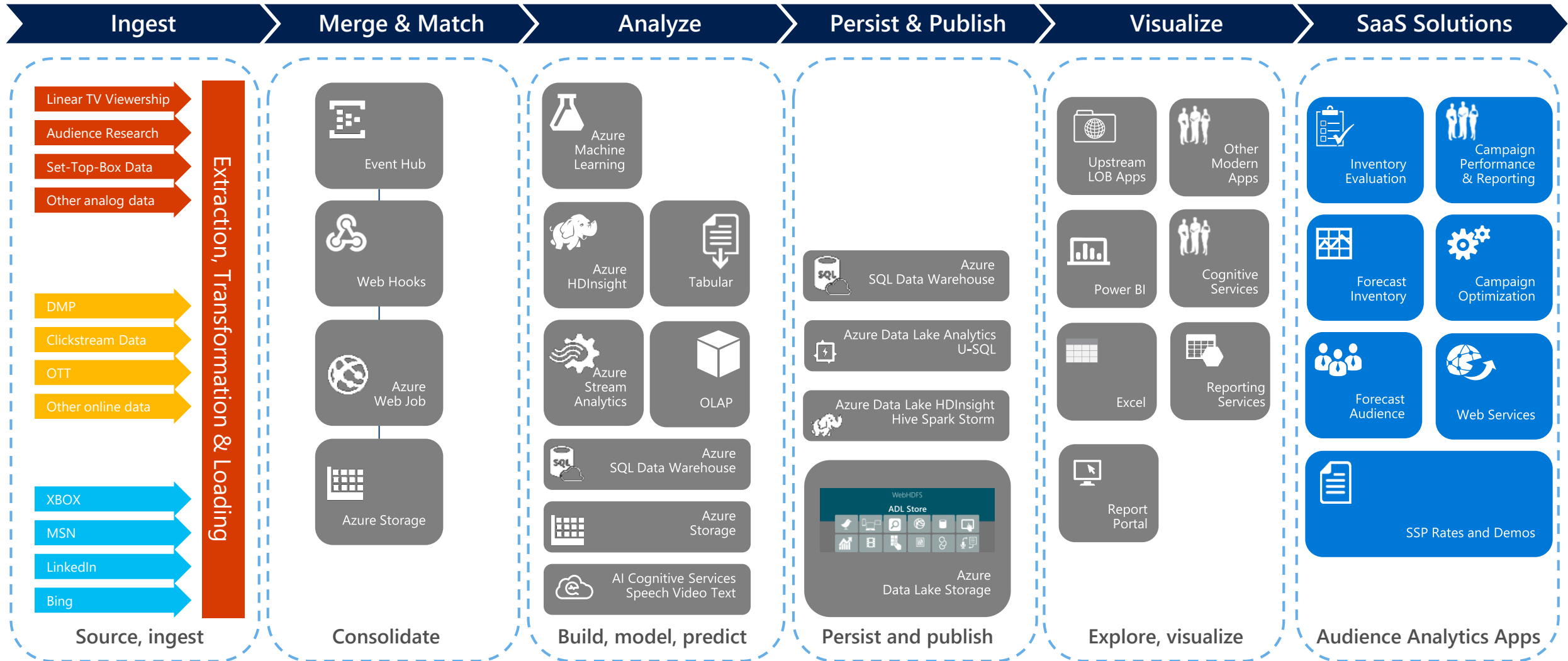


What kind of content should I acquire?



Which communities do my audiences belong to?

Building an Audience Analytics Platform



Azure is the best place for putting AI to work for you

Sophisticated pretrained models
Most comprehensive set of pretrained services



Vision



Speech



Language



Search

Popular frameworks
Open & interoperable



Pytorch



TensorFlow



Keras



Onnx

Productive services
Machine learning at scale



Azure
Databricks

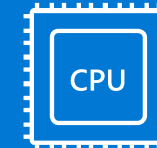


Azure
Machine Learning

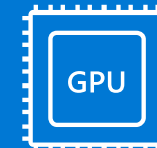


Machine Learning
VMs

Powerful Infrastructure
Most comprehensive
Lowest cost inferencing using FPGAs



CPU



GPU



FPGA

Flexible deployment
From cloud to edge



On-premises



Cloud



Edge

<https://azure.microsoft.com/en-us/services/cognitive-services>

Thank You!



Martin Wahl
Principal Program Manager, Cloud AI & Cognitive Services
Redmond, WA
mawahl@microsoft.com