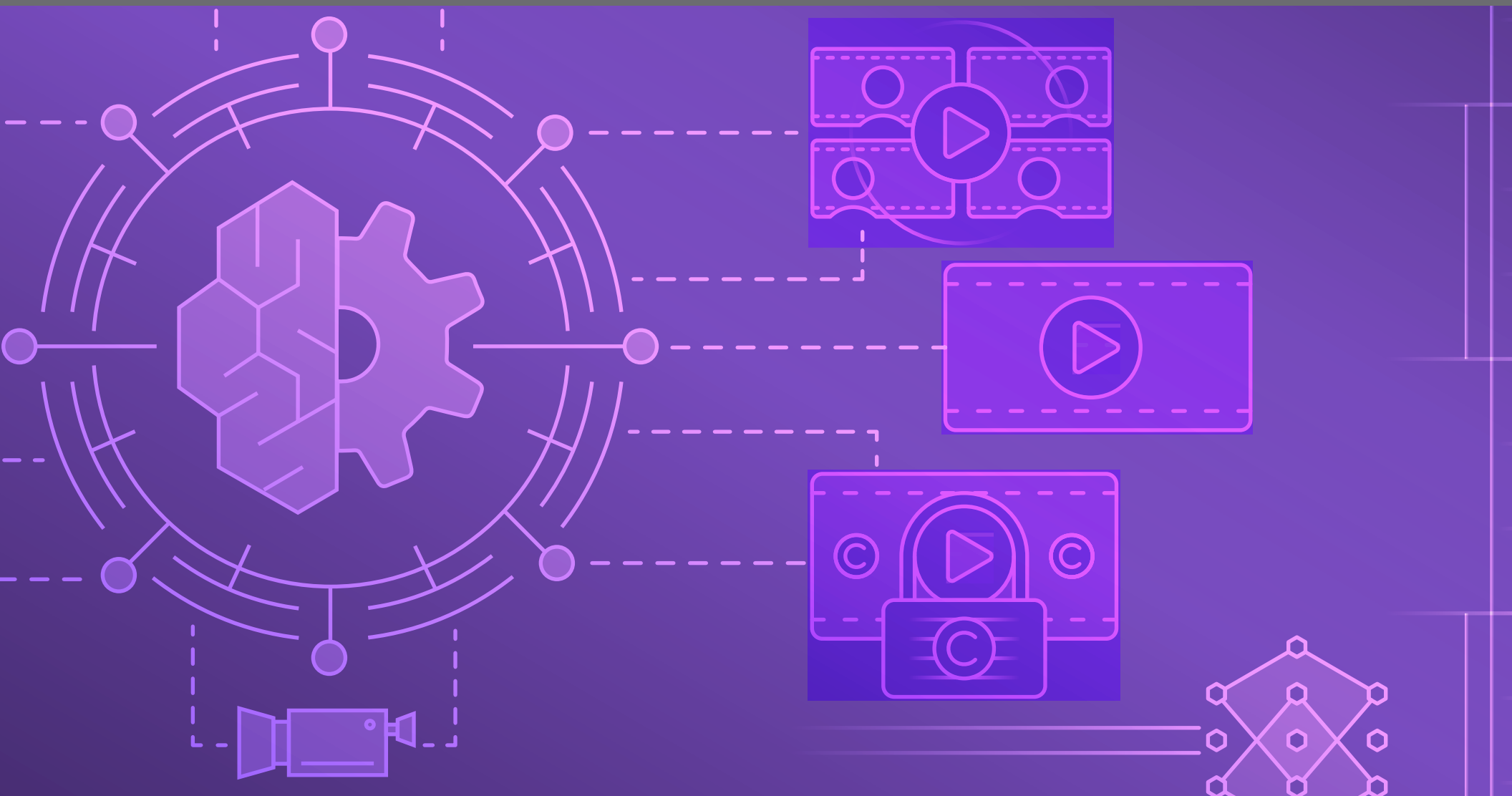
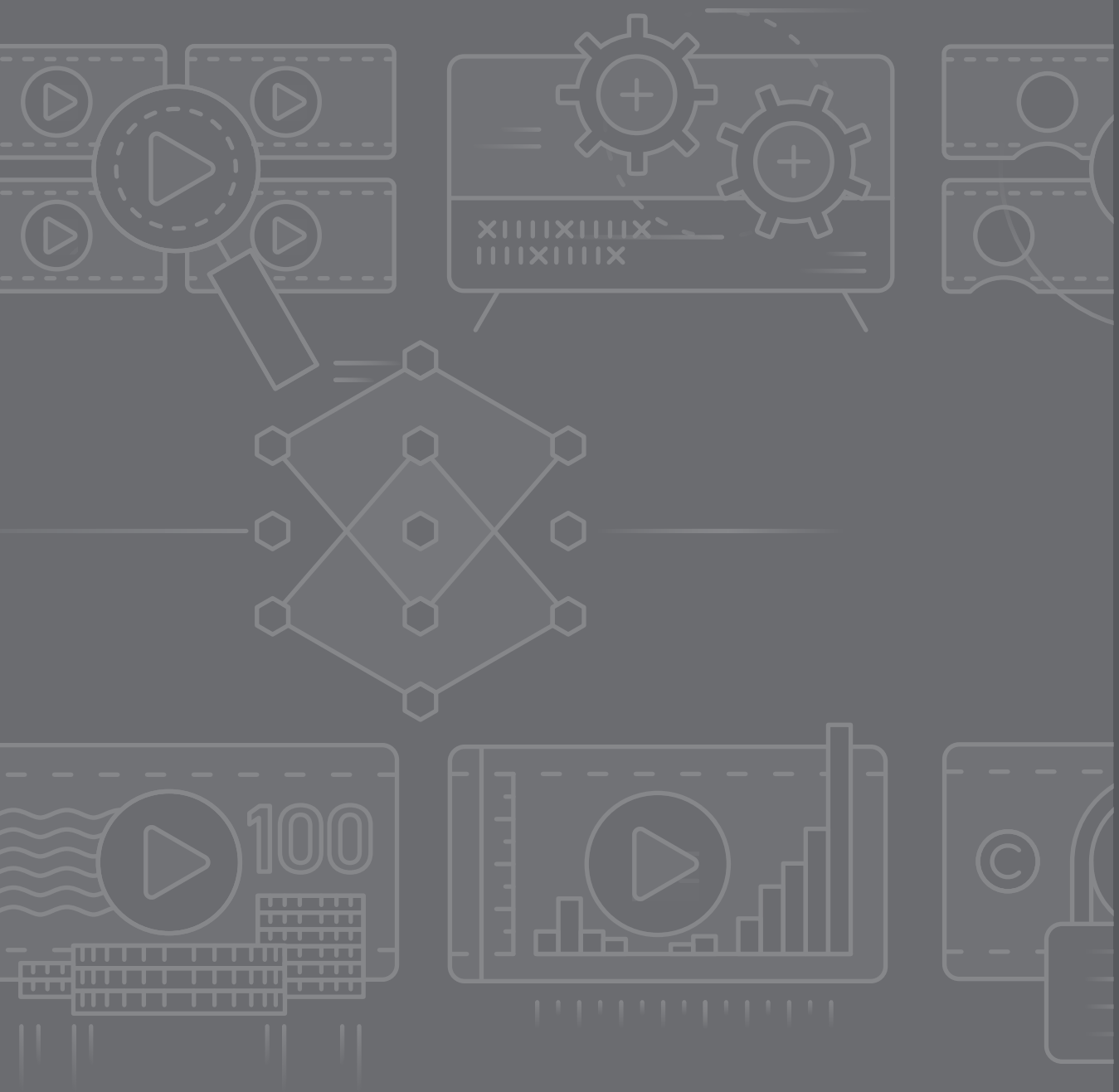
E-book

Building Intelligent Video Workflows with Machine Learning

Cloud-based solutions unlock smarter processes and new user experiences





CONTENTS

01 INTRODUCTION

Introduction

Organizations are adopting cloud-based machine learning solutions at a fast pace. According to Gartner, "By 2019, citizen data scientists will surpass data scientists in the amount of advanced analysis produced. By 2020, more than 40 percent of data science tasks will be automated, resulting in increased productivity and broader usage by citizen data scientists."[i] Additionally, Gartner states, "by 2020, artificial intelligence will create more jobs than it eliminates. AI will create 2.3 million jobs in 2020, while eliminating 1.8 million."[ii]

Organizations looking at ways to incorporate machine learning into their businesses, or to automate and improve existing data science efforts can look to Amazon Web Services (AWS) for solutions that can help. Amazon has leveraged machine learning for twenty years to power its businesses. AWS is now taking Amazon's deep expertise in this space to make sophisticated machine learning platforms accessible to people with little or no background in machine learning or data science.

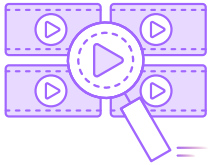
Today, customers are using machine services in greater numbers to prepare data for analysis, build and fine machine learning models, and take advantage of end-user cognitive applications including voice recognition, image and video analysis, providing forecasts and recommendations, and many other intelligent solutions.

This is increasingly true for video providers as well. Whether they work in media and entertainment, enterprise, or the public sector, machine learning holds significant promise for video providers to increase the value of their content and create outstanding experiences for their audiences.

The applications for machine learning in video processing and delivery are vast in number and continuously being developed and refined. In this e-Book, you'll learn about new use cases for machine learning that video providers can benefit from to drive efficiencies, create or enhance offerings, and add value to media assets, regardless of your organization's level of data science expertise.

i Source: *Gartner, 100 Data and Analytics Predictions Through 2021, Douglas Laney, Ankush Jain, 20 June 2017.*
Ii Source: *Gartner Press Release, "Gartner Says By 2020, Artificial Intelligence Will Create More Jobs Than It Eliminates." https://www.gartner.com/newsroom/id/3837763*

02 USE CASES



Searchable Video Archives

With cloud-based machine learning services, video teams can substantially reduce the time and resources spent cataloging, searching, and building assets from their video archive. Machine learning-powered content indexing with metadata generation enables a number of applications with significant real-world benefits.

Many broadcasters must maintain massive archives of video content, often originating from disparate sources and using inconsistent, if any, systems for tagging assets. With machine learning tools, the time-consuming manual labor of tagging content for search can be eliminated. Content libraries are quickly, automatically optimized for fast, accurate search.

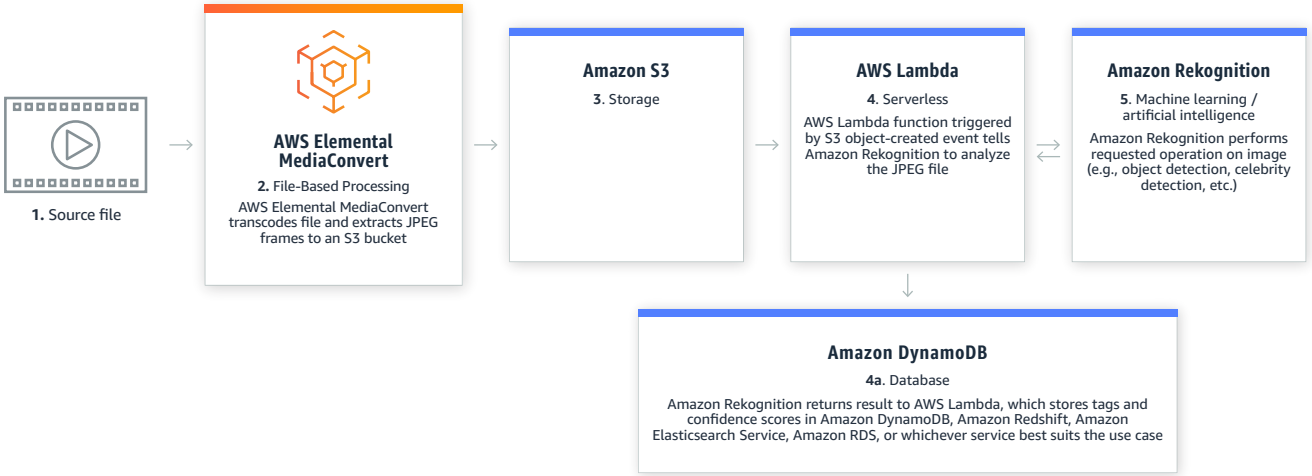


Figure 1 – 5

There are two approaches to indexing content for search that can be applied depending on the length and subject matter of the assets. For very long content, such as parliamentary proceedings which may run several hours per session, simply extracting still image frames from video is the most efficient approach to building a comprehensive index of the content, as seen in Figure 1 above.

Using AWS Elemental MediaConvert, a file-based video transcoding service from AWS and a part of AWS Media Services, users extract still image frames from five-second segments of every video asset in an archive. These still images are furnished to Amazon Rekognition, a service that makes it easy to add visual analysis to applications, for image analysis and extraction. Amazon Rekognition analyzes the still images to create metadata associated with the video content. Videos can be tagged based on objects, celebrities, and other details of the particular scene. The result is a searchable index of metadata that editors and producers can rely on for fast, easy discovery of source content for file footage, clip generation, and more.

For shorter content, such as an hour-long TV show, it is best to process the full video rather than capturing image frames. For example, a content producer who wants to log which actors are in each scene can process a complete show, in file format, using Amazon Rekognition and AWS Elemental MediaConvert to automate the content indexing process. As illustrated in Figure 2, Amazon Rekognition Video analyzes the files and creates metadata that logs when each celebrity is on screen and their location in the shot, with timestamps. Again, a time-consuming manual process is automated, yielding a dependable index of searchable metadata.

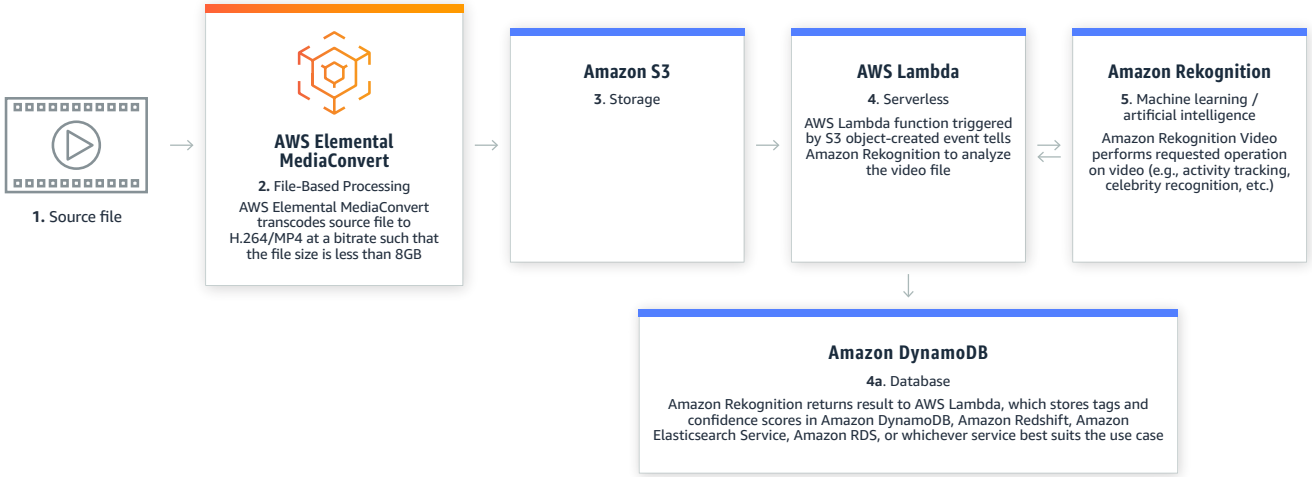


Figure 2 - 5

Automating Video Captions

Caption metadata is essential to making video accessible to all audiences and may be required by law depending on content and region. However, the process of transcribing video assets, then producing and integrating accurate captions in the various formats required to assure availability across different screens and devices, can be costly and slow. The time and expense involved becomes increasingly prohibitive at scale, when generating captions for large volumes of content.

The advent of machine learning tools that can process and analyze video in the cloud gives content providers a powerful, scalable process for automating caption creation. As an example, an online training provider has thousands of hours of video, all of which need captions to meet accessibility requirements set forth by its newest corporate customer. Videos are archived in a variety of file formats, such as MPEG-2 and AVC (H.264).

Using AWS Elemental MediaConvert, the training provider in this example can generate an audio-only rendition of the file for analysis. The audio-only .wav file is provided to Amazon Transcribe, an automatic speech recognition service, which generates a time-stamped transcription of the audio as a JSON file. Then, that file could be accessed by an AWS Lambda function to generate closed captions in each of the required formats, such as WebVtt, SRT, and TTML, which could then be integrated into the complete transcoded file for delivery across different screens and device types. To further enhance captions, a natural language processing service such as Amazon Comprehend can be applied to the transcribed text to identify insights and relationships, thereby providing further context to accompany transcribed dialogue.

As an extension of this workflow, users can also choose to process the transcribed file with Amazon Translate to generate captions in multiple languages, including Arabic, Chinese, French, German, Portuguese, and Spanish. A neural machine translation service, Amazon Translate uses deep learning models to deliver more accurate and more natural sounding translation than traditional statistical and rule-based translation algorithms.

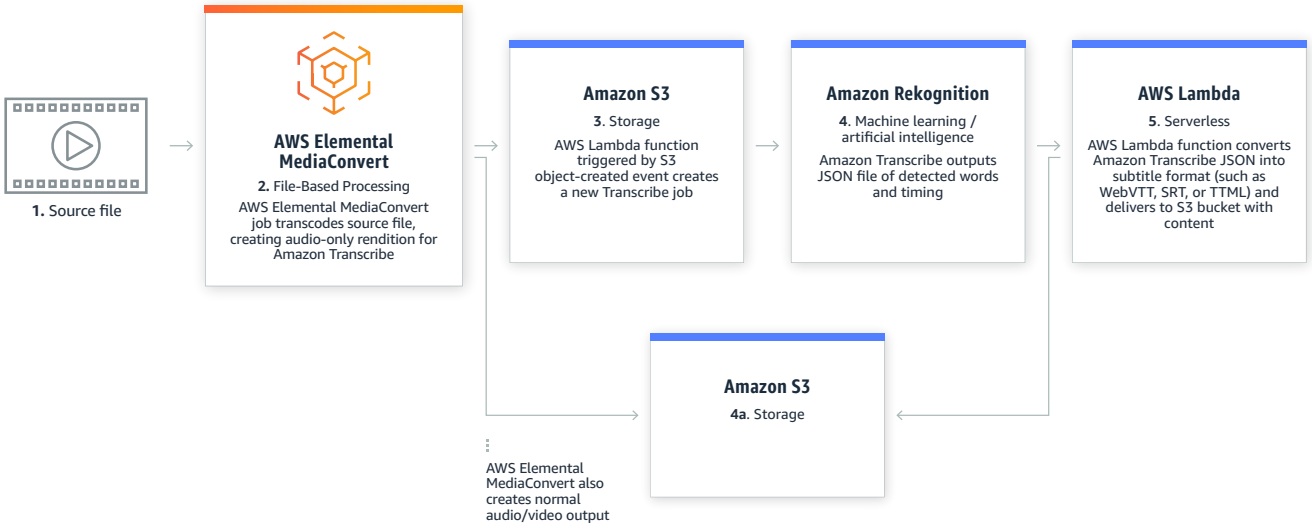


Figure 3 - 5

Video Clip Generation

Social media is a uniquely powerful tool to draw audiences to news and entertainment content. Traditionally, the process of generating and publishing clips has required a manual workflow to identify relevant content from source video, generate time-coded clips, then transcode, package, and distribute those clips for publication on social channels.

This high-touch, multi-step process can be time consuming and result in missed opportunities, particularly for live event broadcasts. Now, machine learning tools can automate key steps of the process to help broadcasters get high-value clips to viewers' screens in near real time.

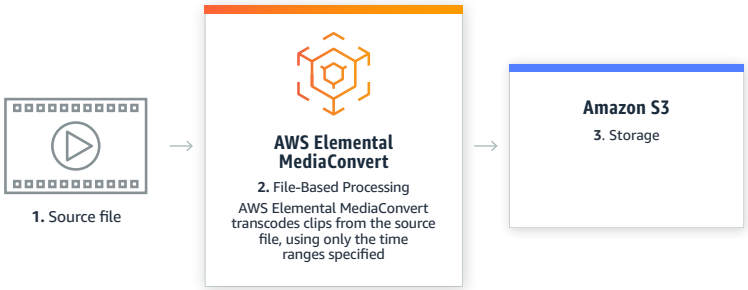
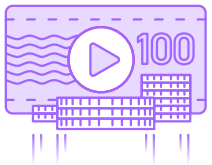


Figure 4 – 5

As an example, a content producer for a weekly cable TV drama would like to create a promotional clip for an upcoming episode with a popular guest star. Using Amazon Rekognition Video's celebrity detection capabilities, the episode's source file is processed to identify all of the scenes in which the celebrity appears. The system then submits the raw video with accompanying time references to AWS Elemental MediaConvert, which selectively transcodes only the segments with the identified star. These segments are delivered to Amazon S3, ready for editors to quickly and easily edit into b-roll and standard show assets to produce a promotional package for broadcast, OTT, and social media distribution.



Personalization and Monetization

To optimize revenue opportunities for streaming video, content providers must equip their infrastructure to furnish advertising that is targeted to individual users and delivered in a way that enhances the viewing experience. In fact, recent research shows seventy-eight percent of consumers say they prefer personalized ads.[i]

With machine learning-enhanced video workflows, content providers can now seamlessly insert personalized advertising based on a variety of factors, such as the type of device being used by the viewer, demographic information about the viewer, or even information about the content being streamed, known as *content-aware advertising insertion*.

For example, an OTT services provider holds rights to a percentage of advertising slots for all of the programming it delivers. It markets that ad space to advertising networks, which provide advertising via ad decision servers that determine which ads to serve based on the information made available to it by the content provider or video player; typically demographic or geographic data about the viewer. It markets that ad space to advertising networks, which provide a vast pool of commercials via ad decision servers from which the content provider can draw. Using Amazon Rekognition video analysis and Amazon Transcribe speech recognition, the provider uses machine learning to continuously feed a database of keywords that reflect the content on screen. By applying machine learning models built with a service such as Amazon Sagemaker, it becomes possible to create an intelligent ad recommendation platform powered by machine learning.

To illustrate how this works, consider a 30-minute program in which the scene previous to an ad break heavily features house cats. The video workflow signals AWS Elemental MediaTailor, a cloud-based content personalization and monetization service, to transcode relevant advertising on-the-fly and insert it into the stream. The service calls on the ad decision server to deliver ads most closely associated with the content as dictated by the ad recommendation platform. As a result, a 15-second spot for a new brand of

i Source: *Adobe Digital Insights Advertising Demand Report 2016: North America*

premium cat food is inserted for the next ad avail. Following that, a 10-second commercial for pet adoption services.

Over time, the content provider uses closed-loop metrics supported by AWS Elemental MediaTailor to validate that its content-aware advertising outperforms other methods of ad delivery. Using rich measurement data, the provider is able to justify increased fees to advertisers for its commercial space based on its proven ability to engage audiences through its advertising.

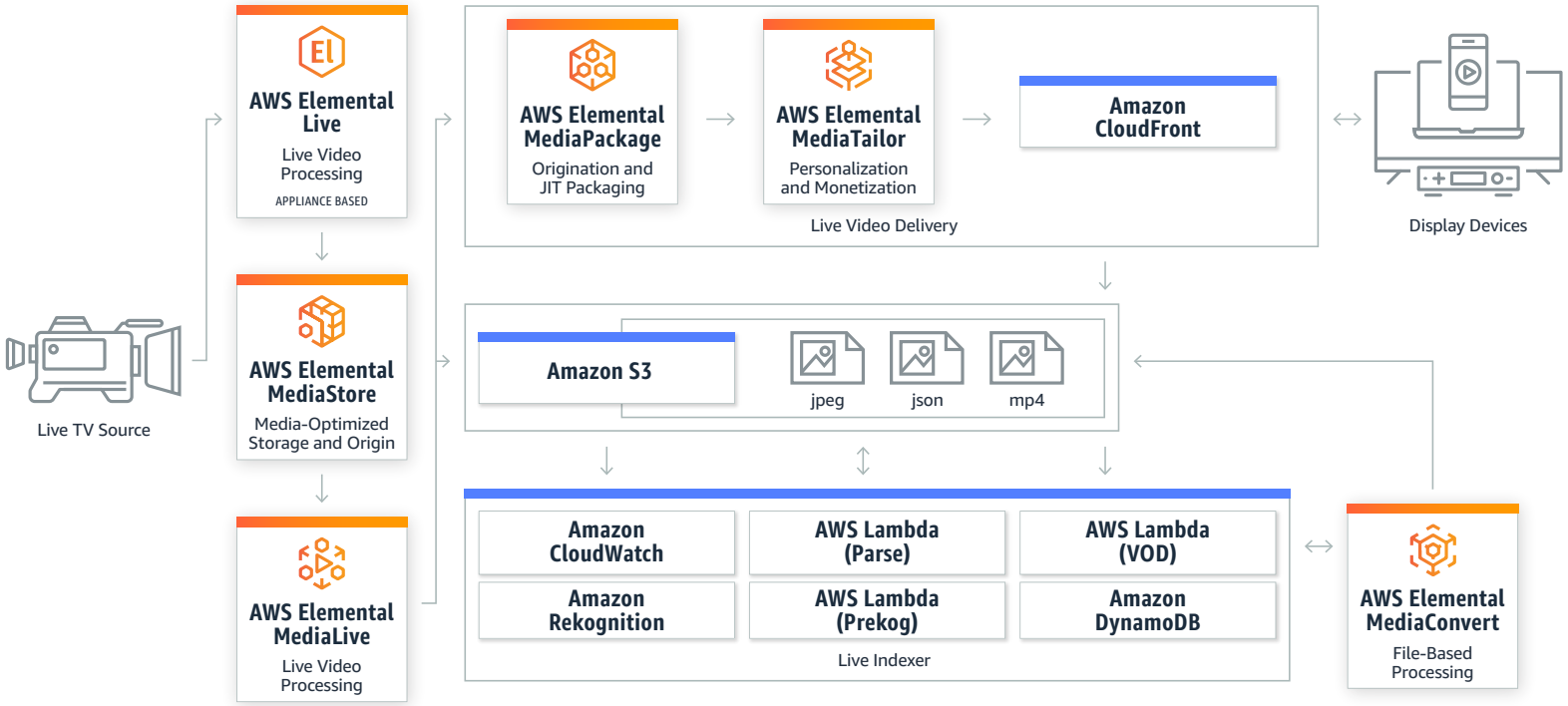


Figure 5 - 5

Analytics and Measurement

Today's video workflows have the ability to measure and report vast amounts of information about live and VOD streams and the infrastructure behind them. Data related to the performance of individual components, key processes, and complete workflows can be measured and used for real-time notifications or long-term analysis. By applying machine learning systems, content providers can uncover new paths to optimize every aspect of the video workflow, including workflow performance, use of network resources, monetization results, and much more.

For example, an OTT service has secured rights to live stream games for a major international sports league. The games are also broadcast by a different rights-holder, using a separate workflow; the only elements of the workflow shared between the two organizations are the live feeds from on-location satellite trucks. Heavy viewership of the broadcast channels means that viewers are describing the action online and via social channels, play-by-play, in real time.

To present a worthy alternative to the broadcast service as well as offering a valuable second-screen experience, the OTT provider sets a goal to deliver live video as quickly as, if not more quickly than, the broadcast infrastructure. This requires careful engineering of the OTT video workflow to reduce latency at every step, supported by robust real-time measurement to enable continuous optimization across the workflow.

Using machine learning services, the OTT provider is able to integrate real-time analysis of performance across the video workflow, mining the constant flow of data to identify sources of latency as they arise and invoking adjustments, such as scaling distribution resources, optimizing load-balancing, or redirecting CDN traffic across different paths or regions as required to maintain minimum latency. The provider is able to preserve faster-than-broadcast latency and keep its viewers ahead of the action on the field of play.

Copyright Protection and Content Security

When video professionals think about security in the cloud, the first thing they think about are defending against potential incursions by malicious actors. However, many content providers will cite simple human error as one of the most significant threats to the security of content, regardless of where content is stored.

Imagine a storage volume containing footage from an unreleased Hollywood blockbuster which an editor at a post-production house has mistakenly exposed to Internet search and you can understand why avoiding such a possibility is a high priority for content providers.

One forward-looking application for machine learning is an intelligent tool that scans online video for unreleased content. Such a tool can use a reference database of stills or clips from unreleased video and scour the potential hosts on the public Internet (or "dark web") for leaked or misappropriated content. The tool would identify online content that matches the database and immediately flag the organization with details on the location and nature of any copyrighted content it locates.

A similar application involves identifying content that has been deliberately altered to avoid copyright enforcement. In such cases, individuals are posting copyrighted content to video sharing websites with subtle alterations, such as obscured watermarks or frame rates that are one frame-per-second slower than the original, that may avoid automated content filtering approaches. In the future, with machine learning, this type of infringing content can nevertheless be discovered. To do so, machine learning tools would again rely on a reference database of content. In this case, that content could be text; specifically, scripts of copyrighted content. Using automated transcription tools, video sharing sites could be regularly scanned for video whose scripted audio matches a script in the database. Regardless of how the video may be altered, this approach would provide an audio-only option to ensure copyright enforcement for valuable content.

This e-book addresses a handful of the practical applications for machine learning services that can enhance video workflows. Whether on a mission to Mars or a path to operational excellence, every organization that relies on video can now look to machine learning as a deep, untapped resource from which to drive efficiencies, create new services, enhance current offerings, and add value to their content. Video providers are encouraged to explore for themselves the array of easy-to-use systems and self-service tools available to start building their own machine learning solutions.

AWS offers several tools and information resources to help content providers begin to incorporate the power of machine learning in their video workflows today:

AWS Media Analysis Solution

Featured Machine Learning Partner Solutions

AWS Machine Learning Blog

AWS Media Blog

To get the latest industry insights, bookmark the AWS Elemental Blog, and follow us on your favorite social channel for more client stories, events and news.

About AWS Elemental

AWS Elemental, an Amazon Web Services company, combines deep video expertise with the power of the AWS cloud. Solutions from AWS Elemental allow broadcast TV and multiscreen video to be customized, originated and monetized at global scale. Flexible, software-based video processing and delivery gives global media franchises, pay TV operators, content programmers, broadcasters, government agencies and enterprise customers the ability to deliver highly differentiated viewing experiences and the freedom to focus on what matters: transforming ideas into compelling content that captivates viewers.