

# Controlled Ascent



Some practical clarity on media migration from LTO to the Cloud  
for Broadcast & Production

kumulate

June 2020

Continued advances in storage technology, and a potentially permanent shift in operational environments and processes, has resulted in a majority of media & entertainment organizations considering moving their video asset repositories, currently stored in on-premise LTO tape libraries, either partially or entirely into cloud-based storage. Providers such as Amazon Web Services, Microsoft, Google, Wasabi and other public cloud hosts are becoming ever-important players in M&E video storage.

Storage ecosystems (comprising short term disk storage, and mid- and long-term storage) underpin the entire media workflow infrastructure. As such, any reconstruction needs to be carefully planned and evaluated, to ensure that the correct ecosystem is designed to match and optimize storage for specific types of content, its lifecycle, and usage.

This white paper has been developed by Masstech to examine and discuss some of the most major considerations that apply to any large-scale migration of content, and we hope that if you're considering a migration from LTO to cloud, you'll find it a useful reference.

---

This paper was prepared by multiple Masstech technical staff. Special thanks to Mike Palmer, who prepared much of this document in his role as CTO at Masstech; and to the IABM, whose research has been quoted, and without whose constant efforts reference works such as this would not be possible.

You can discover more details about content migrations, and about Masstech's intelligent cloud and hybrid content management platform Kumulate, at [www.masstech.com](http://www.masstech.com), and if you have any questions at all please reach out to us via email at [info@masstech.com](mailto:info@masstech.com).

## Contents

- Introduction ..... 4
- Storage ecosystems - types of deployment..... 4
- Which model is the right model for you?..... 7
- Considerations ..... 7
- ISM: The importance of media management before, during & after ..... 7
- How much content and storage do you have? ..... 9
- Understand your on-prem hardware and networks ..... 10
- Determine how much bandwidth is available between you and the cloud..... 11
- Understanding your cloud options ..... 12
- Cost and Performance Modeling of OpEx Public Clouds and CapEx Private Clouds... 13
- During the migration ..... 15
- ISMs and post-migration data independence ..... 16
- Data integrity and evolution management ..... 16
- Turning a chore into an opportunity ..... 18
- Conclusions and Takeaways ..... 21



## Introduction

Most organizations have or will consider moving some or all of their on-prem content to private or public cloud storage, and in doing so will begin a deep dive into ever-increasing levels of inter-related detail regarding cost, features, integration, performance and time. Though the decision to migrate content from on-prem LTO to the cloud, and indeed whether to work in one of several hybrid environments is different for everyone, every organization will need to understand and calculate for themselves the site-specific variables and options that will lead to their unique business decision.

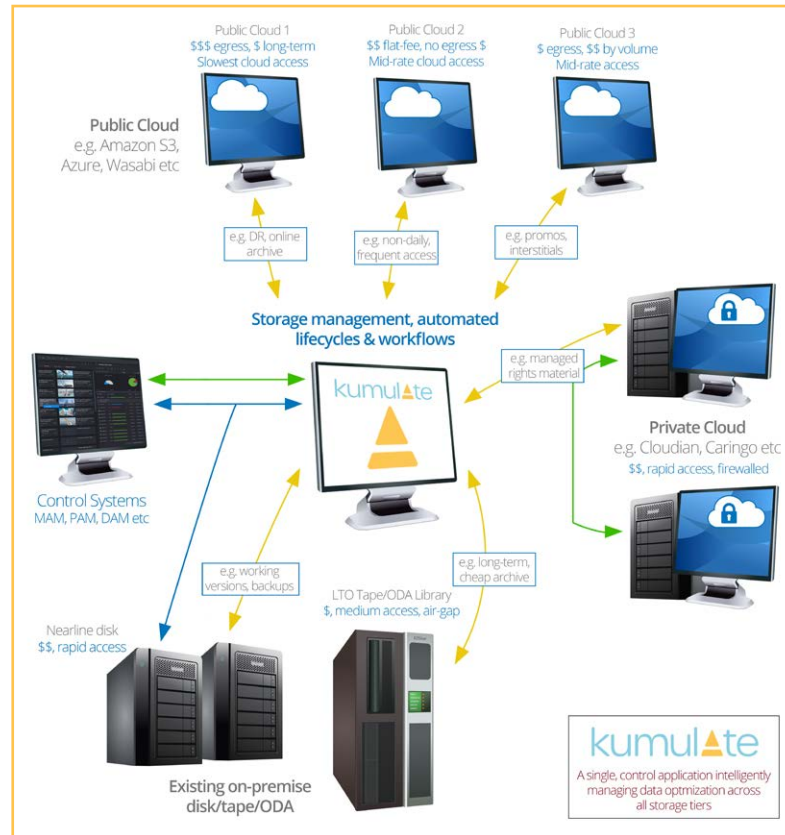
The business drivers for content migrations are diverse. A perception that cloud is less expensive is generally a prime consideration, but the potential efficiencies of locating content physically closer to scalable compute services, such as distribution, transcode and AI services are other valuable benefits. However, the technical models are complex, even when looking at the very high-level permutations of specific hybrid cloud operations.

The purpose of this document is to familiarize you with the processes, variables and options you'll need to consider, and briefly describe their interaction. To start with, let's take a look at the different types of deployment.

## Storage ecosystems - types of deployment

- 1) Public Cloud only – Public cloud storage is immediately attractive for start-ups and green-field deployments where availability, responsive performance and distribution are immediate attractions, but storage of very large amounts of content for a very long time is not a requirement. Types of business that may be attracted to exclusive use of public-cloud storage are production houses and studios working on creative episodics, special effects production and long-form cinema.
- 2) Private Cloud only – Private clouds are generally thought of as the realm of enterprises who generate a very large volume of content and have the expertise and resources to build private storage which

*Tape, disk, private cloud, public cloud, hybrid, multi-cloud - many deployments are possible. Choosing the right combination requires a lot of planning*



approaches the scale, but can exceed the cost efficiency, of public cloud providers. Types of businesses that might be attracted to private clouds would be large news organizations and organizations that have very strong needs for very long-term preservation but low throughput. Some customers also have business and contractual needs to maintain direct physical control of their assets.

- 3) Public and Private Cloud together – Customers who have a need to keep direct control of assets, for security or cost reasons, may have the need to burst distribution volume through a public cloud. Public/Private hybrids are also very attractive for disaster recovery, as well as for providing means for an enterprise to integrate content and distribution in a controlled manner with external business and service providers.
- 4) Public Cloud or Private Cloud combined with On-Prem storage – Any migration that requires uninterrupted access to content will need at least some period of hybrid operation, whether using a Public or Private cloud. Beyond that, a cloud/on-prem hybrid can be used to provide cloud proxies while keeping full bit-rate assets local, or the reverse – full bit-rate in the cloud and proxies on-

prem. Either architecture can be completely valid depending on the business requirements.

- 5) Multiple Public, Private and On-Prem storage together – In this model content is managed in the location with the best combined attributes for immediate or next use, without much regard for pedantic technology labels. Instead, a smart multi-domain storage management system makes a number of storage locations available for use and automatically and dynamically moves content, based on associated metadata, to the location that has the best commercial and technical attributes for the next business operation.

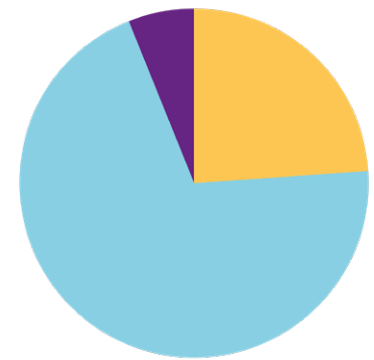
Media tech preferences  
data courtesy of IABM

---

“... Media companies have so far shown a preference for hybrid cloud solutions with the adoption of public cloud depending on the use case. The cloud has brought a lot of complexity as well which is why a major focus going forward is on enabling multi-cloud workflows, to avoid any vendor lock-in”

IABM: IBC 2019 Special Report

---



- Prefer a complete system from an end-to-end provider
- Prefer to create a 'best-of-breed solution'
- Don't know

## Which model is the right model for you?

The suitability of each of these deployment models depends on the type of content you have and your specific business model. A deeper dive is required to make an appropriate choice.

### Considerations

The technical cost and performance modeling needed to make a carefully considered decision go far beyond the simple cost per month to store a TB of content, or the capital and operating cost required to build and operate a private cloud. Questions must be considered regarding content normalization, content and metadata storage, de-duplication, consolidation, proxies, AI analysis and metadata services. All of this is in addition to cloud ingress and egress costs.

In addition to cloud considerations, a thorough examination must be made of the existing local LTO storage. Decisions must be made as to whether to keep on-prem LTO, and whether to use that existing capex investment, the number of LTO generations presently supported, the number of LTO drives available, and most critically how to provide uninterrupted service to existing users and services while any migration is in process. If users need uninterrupted access to content during the migration, this means that, at least for a time, a hybrid model must be considered.

### ISM:

## The importance of media management before, during & after

A spreadsheet to model and rationalize the relevant variables quickly becomes lengthy, even before we can begin to knit it together with appropriate calculations. However, the spreadsheet only helps us make a decision. We need another tool, something more than the typical HSM, to securely conduct, manage and monitor the progress of the migration, and critically, to provide a seamless bridge for all integrated systems to access content as it exists during and after the migration both on-prem and in the cloud.

This is where Intelligent Storage Management platform, (such as Masstech's Kumulate) enters the fray. We say "intelligent storage management" because this is the next



generation of HSM, going beyond simple management of multiple tape and disk storage systems, to manage large amounts of content and metadata within a single name space, simultaneously integrated with any number of public and private cloud object storage services, conditional and concatenated workflows and decision trees, wrap, re-wrap and transcode services if needed, and active monitoring. The purpose of the intelligent storage management is not only to efficiently manage multiple storage locations and buffer MAMs, PAMs and distribution systems from constantly evolving storage technology, but crucially to protect the fundamental integrity of content over time.

---

“The power of the ISM to your organization is its ability to apply your specific business rules to a mesh of constantly evolving technology and services. The value of the ISM is the continuity it provides to users and integrated systems in these quickly changing times, allowing uninterrupted and deep access to both content and metadata”

---





## How much content and storage do you have?

Determine how many assets you currently have on premises and their specific file sizes. Though it is counterintuitive, the bulk transfer of small files can be much less efficient than larger files. This is because read, transfer and write processes used in the migration commonly involve a fixed amount of process overhead per file, regardless of file size.

An exception to this rule is the **use of file wrappers** which create a larger file by aggregation of large numbers of smaller files. Wrappers increase efficiency by allowing many files to be read, transferred and written with the overhead equivalent to only a single file. Examples of common wrappers are, ZIP, TAR, AXF ,etc. We will have more to say about how the use of wrappers in the cloud can reduce storage costs later.

**De-duplicate your assets** unless you're absolutely sure nearly all of your content is unique. Migrating only unique content will decrease cost and increase speed and efficiency of any migration. This can be done during the migration, but if done in advance it will allow better planning, including planning for the amount and cost of storage in the destination cloud(s). A quick and first method usually involves searching for files that have the same name and file size. A more accurate, but time-consuming method is to search for content with the same file size and hash. The second method will identify exact bit-for-bit duplicates but will not identify content that is creatively or editorially a duplicate but stored in a file of a different format or bit rate. Keep in mind that hash generation is time consuming. If hashes have not been created at the time of ingest or transformation, generating new hashes from all content will take some time.

**Count the number of physical LTO tapes**, noting their LTO versions and whether they are physically available to a tape robot, or whether they are located on a shelf. Counting the number of tapes will only give you a rough idea of the number of TB of content to be transferred, as tapes are rarely completely filled. Also, if any content was deleted from any LTO tape, it is likely the content was simply marked as deleted and in fact still takes up space on the tape.

## Understand your on-prem hardware and networks.

Take note of the available LTO drives and their versions. We assume there is an appropriate combination of drives to support the different versions of LTO that might be stored on-prem. Keep in mind that not all LTO drives can read two generations backwards - this depends on the specific generation of the LTO drive. Also, different generations of LTO drives read and transfer data at significantly different rates. If the transfer rate had kept pace with increasing data density, transfer speed of LTO-8 would be around 20Gbps. However, in actuality, LTO-8's transfer speed is limited to only about 3GBps.

One of the more difficult to obtain pieces of information you'll need is the daily duty-cycle on each drive. This is the amount of time each day in which a drive is actively reading or writing a tape under normal operation. The amount of time that drives are inactive is the amount of time they are available to read content for the migration.

Since most customers did not over-design their on-prem systems, they may find the amount of normally inactive drive time available to the migration process may be surprisingly low. We will talk more of this later as for some it might prove to be the largest limiting factor in the migration process.

---

“Most broadcasters [say] that data is very difficult to get right as their legacy architectures has allocated it in different silos...Correcting this is key in deploying solutions in an effective way”

IABM: IBC Special Report 2019

---

## Determine how much bandwidth is available between you and the cloud

This is a function of both how much bandwidth you can find from a vendor and how much you can afford to purchase. Some public cloud vendors offer direct connections which can significantly boost performance.

Sometimes the fastest connection to the cloud is not an internet or WAN connection but a NAS connected directly to your Kumulate ISM platform onto which content can be quickly copied. Once the NAS is full it can be physically shipped to the cloud where it can be copied into permanent storage. A transportable NAS will typically offer greater effective bandwidth than can be purchased through a network connection over the public internet, even with the additional overhead of redundant copies and shipping delays.

However, many sites may find available bandwidth from existing LTO drives will be a critically limiting factor (inactive drive time x LTO transfer speed) and therefore will not prohibit the full use of large internet or private network pipes. In fact, in many cases, the amount of available bandwidth from existing LTO drives can be so low that working with a NAS can be no faster than moving the data over existing internet connections.

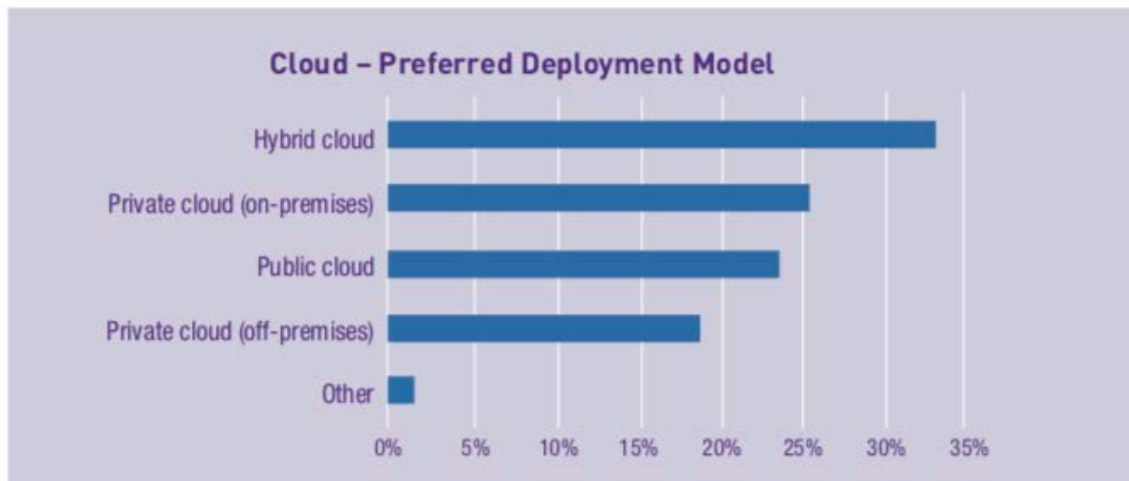
Four solutions to the challenge of limited LTO bandwidth present themselves.

- 1) The first is to simply have end-users stop driving the system, thus freeing the full bandwidth of all available drives. This is likely not an option for most sites as daily operations have to continue and content is sent to tape daily.
- 2) The second is to rent or buy additional LTO drives, up to the maximum physically possible to install in the existing library. If your LTO tapes include more than two successive generations, keep tape/drive compatibility in mind as you add drives.
- 3) The third option is to rent or buy a small LTO library with a completely separate set of drives, and manually shuffle tapes through it. In fact, if the selection and sequence of content is thoughtfully made, downstream copies to new systems can subsequently be achieved more quickly and efficiently.
- 4) The fourth option is to simply run the migration processes at a lower priority

than user-driven workflows and make the most of all available drive time and bandwidth. Essentially, this approach ensures the LTO migration takes place when the system is not otherwise in use or during down time, regardless of how short or long these periods might be. The natural consequence of this of course is that the entire migration is likely to take longer than if tape resources were specifically dedicated to it.

## Understanding your cloud options

Cloud storage solutions and cloud compute options are not created equally. This is a good thing. Some cloud services are optimized for lowest cost, long-term storage. Others work better for short-term storage of content intended for distribution. Some offer various degrees of redundancy to improve the durability of data both infrastructurally and geographically. Others provide the benefit of proximity of data to scalable compute and AI services. You may well find your best solution to be a combination of storage and compute provided by multiple clouds.



While hybrid cloud solutions have become increasingly attractive, all clouds have different price and business models. This provides the flexibility to assemble best-of-breed solution, and the capacity to change components as technology and services evolve, but it also makes analysis of technical and business models more difficult.

## Cost and Performance Modeling of OpEx Public Clouds and CapEx Private Clouds

In addition to understanding how much on-prem content you wish to migrate to the cloud and what optional processes you wish to apply, you'll also need to know how much content you're putting into and pulling out of storage on a monthly basis. You'll need this information for two reasons: 1 to compare the cost as well as specific features of cloud storage across a wide range of public and private architectures and vendors; 2 to factor the amount of content into bandwidth calculations needed to model performance and cost.

Performance models will consider the number and size of files, number of tapes, transfer speed of LTO drives (specific to LTO generations in use), overhead of physical tape handling, availability of drives exclusive of user-driven workflows, network bandwidth, efficiency of network object transfer processes (eg S3), optional network acceleration, and optional processes.

Pricing models will take into account the output of the performance model, in addition to cloud storage costs, potential ingress and egress costs, network or NAS costs, and potential costs to temporarily add more LTO drives. Some sites will simply look to place a copy of content in a cloud while continuing to use on-prem storage, either as primary or secondary source. These customers can take advantage of a "relaxed" migration/copy that makes use of drives as they are available and pulls

*Most storage providers offer a calculation service on their web site. You need to be aware of all of your requirements to get an accurate picture of cost, and if you're adding in extra services such as AI/ML make sure you're keeping comparisons on a like-for-like basis*

The screenshot displays the AWS Pricing Calculator interface. The main heading is "AWS Pricing Calculator Estimate the cost for your architecture solution." Below this, there are sections for "How it works", "Benefits and features", and "Compare how Wasabi is cheaper than". The "Compare" section shows a bar chart comparing storage costs for Wasabi, Amazon S3, Microsoft Azure, and Google Cloud. The chart indicates that Wasabi is significantly cheaper than the other services. The interface also includes a "Create an estimate" button, a "Getting started" section, and a "More resources" section. A diagram on the right side of the interface illustrates a modern data warehouse architecture.

Example calculators from AWS, Microsoft Azure and Wasabi show some of the levels of complexity involved in the calculations.



content from the best storage location. Other sites will want a fast and full migration to the cloud, ending in decommissioning of the on-prem storage. For these customers the pricing model will be very important as it can show how to minimize the incremental operational costs during the migration period itself, for instance by adding drives, NAS or additional network bandwidth.

The ability to easily iterate through multiple versions of performance and price models is important in order to handle the inevitable “What if” permutations. For instance, some customers may be more sensitive to storage cost than to maintaining the highest quality. How would reducing the bitrate from 35Mbps to 8Mbps affect both storage cost and time to perform the migration? Or another variation on this in which 25% of the content is maintained at 35Mbps. How much faster could I complete the migration if I added an additional LTO-6 drive? How would this affect the amount of network bandwidth I need? Would a NAS transfer be more efficient?





## During the migration

The ISM platform should monitor actual performance, compare this against the modelled prediction, and alert on conditions outside the model, such as those that would cause delays or excursions from the cost budget.

Where monitoring indicates a failed read or write operation – for instance if hashes do not match after a copy or move or a copy operation fails outright – the operation must be retried. The time required for these additional actions and the frequency of these events must be taken into account.

Throughout the operation, MAMs, PAMs, distribution and other systems continue to access the content through the Kumulate ISM without knowledge of the migration, its status or the changing location of the assets. The Kumulate ISM is responsible for reading/writing content from the appropriate cloud/technology/location.

---

“...media companies have not ceased to show their preference for best-of-breed solutions, even in a cloud world. This has led to a significant rise in partnerships between specialized media technology suppliers and cloud service providers. This has also evolved into a requirement for software integration and multi-cloud workflows, in the most advanced environments. With the advent of these technologies, a new layer of complexity has also been added to media companies’ operations. This rising complexity in media companies’ operations has led to closer relationships with their suppliers.”

IABM: IBC 2019 Special Report

---

## ISMs and post-migration data independence

Even though the content has moved to new cloud locations, the Kumulate or other ISM continues to provide uninterrupted and transparent access to the content to attached MAM, PAM, distribution and other systems via established API integration. From the user's perspective, operation continues as normal even though the underlying storage locations and technologies have changed.

The benefit of using the ISM during and after a migration is that no matter how many formats, wrappers, storage technologies and locations are in use, any number of MAM, PAMs or other systems can interact with normalized content and metadata through a simple object interface without requiring any knowledge of the details of the underlying storage

Your new cloud storage will provide easier and true open access to metadata. If you've chosen to store existing or newly enriched metadata with assets, any system integrating with the Kumulate ISM will have equal access to this data and associated assets. This is known as "Metadata Independence." Because this metadata is wrapped with each asset, external processes can be given read-only access to cloud storage to read this content as a "Media Lake," which can be useful for deep analysis and advanced AI training. Systems accessing the data through the Kumulate ISM need not have any knowledge of individual wrapping formats, as normalized and discrete media and metadata can be retrieved through the API independently.

### Data integrity and evolution management

Consider the need for long term preservation. Placing content in a public cloud doesn't mean that migrations have stopped – they've just moved. The cloud provider will need to migrate files from older media and technologies to newer generations as they become available. With every move comes the potential for data corruption. Cloud technologies include hashes at the block and object level and will make corrections when they find them.

---

"55% of companies say security is an important barrier to cloud adoption"

IABM: IBC Special Report 2019

---

Each cloud service does this in a slightly different way. However, most customers also want an independent method to verify these underlying error protection and correction schemes are working. The Kumulate ISM will regularly check the integrity of files in all cloud object stores against hashes independently stored in the Kumulate ISM. Again, hashes maintained by the ISM are the only way to ensure the content you pull from the cloud in the future is an actual bit-for-bit match with the content you put into the cloud today.

One of the most critical continuing functions of the ISM is management of technology change and evolution. Cloud services and the technologies in use today are guaranteed to change over time. With this change will come changes to business models, service levels and pricing, both by customers and storage providers. Customers require a trusted method to seamlessly and automatically move and maintain multiple copies of content in multiple locations, including multiple clouds. The Kumulate ISM provides customers the critical ability to avoid vendor-lock, even if they don't expect to exercise it. The Kumulate ISM is able to maintain any number of copies of content, in any number of cloud locations, for any purpose, independent of management systems or other control layers.



## Turning a chore into an opportunity

Moving entire or even partial repositories of content can be a daunting task, that you may envisage occupies significant staffing resource. As we discussed previously, with the right ISM migrations don't have to be the pain-points in the way they have historically been viewed. But when moving content, there are a number of processes you can integrate that not only result in your assets being in the new, more efficient location, but that content is now safer, validated, and has a great deal more value in the form of enriched metadata.

So, when you're planning the project, ask yourself this question: During the migration what can we do to increase value and discoverability of our assets?

For many, the amount of time and expense required to read and thoroughly analyze content in deep storage might be a concern. However, a migration presents a rare opportunity to analyze and manipulate content without this overhead as the content is in flight between on-prem and cloud storage. Here are some examples:

- 1) **Derive file hashes, if you don't yet have them.** Not only can hashes be used to immediately identify and eliminate duplicate content, they also provide the only way you can prove to yourself that what you pull out of the cloud at a later date is exactly what you put into it today.
- 2) **Wrap edit projects and DPX sequences.** As mentioned earlier, grouping and encapsulating any number of closely related files in a single file wrapper can decrease storage costs and increase both ingress and egress transfer efficiency.
- 3) **Associate descriptive metadata with the content.** This has numerous benefits, including providing a second location of that metadata independent of and as a backup to a MAM. This practice facilitates business continuity contingencies, long term preservation and formation of an open-access data-lake.
- 4) **Transcode and/or re-wrap assets.** This can include simple normalization, creation of new versions for distribution, DR or alternate storage, or reduction in bit rate (only where appropriate) to reduce future storage costs.
- 5) **Write content to multiple locations.** Different types of storage locations offer



their own unique architectures and benefits, allowing users to match workflow and purpose with the properties of the best storage location and technology.

- 6) **Generate proxies (if you don't have them already).** In addition to providing immediate and economical access for user preview and basic manipulation, proxies are increasingly used as input for AI analysis and metadata generation/enrichment. In some cases, proxies actually work better and at less cost for this purpose than full bit-rate and full resolution versions. If you don't yet have proxies and take the migration as an opportunity to generate them, the cost and speed of later analyzing these assets on demand will be greatly reduced.
  
- 7) **Generate a time aligned speech-to-text (STT) transcript.** Essentially, this is a caption file which includes time markers. Not only does this provide a useful source of metadata to search, but the markers allow users to jump directly to the point in the asset where a search match was found. STT transcripts are also considered to be the foundational layer of TAM (time aligned metadata) to which other AI generated layers can be added (see below).

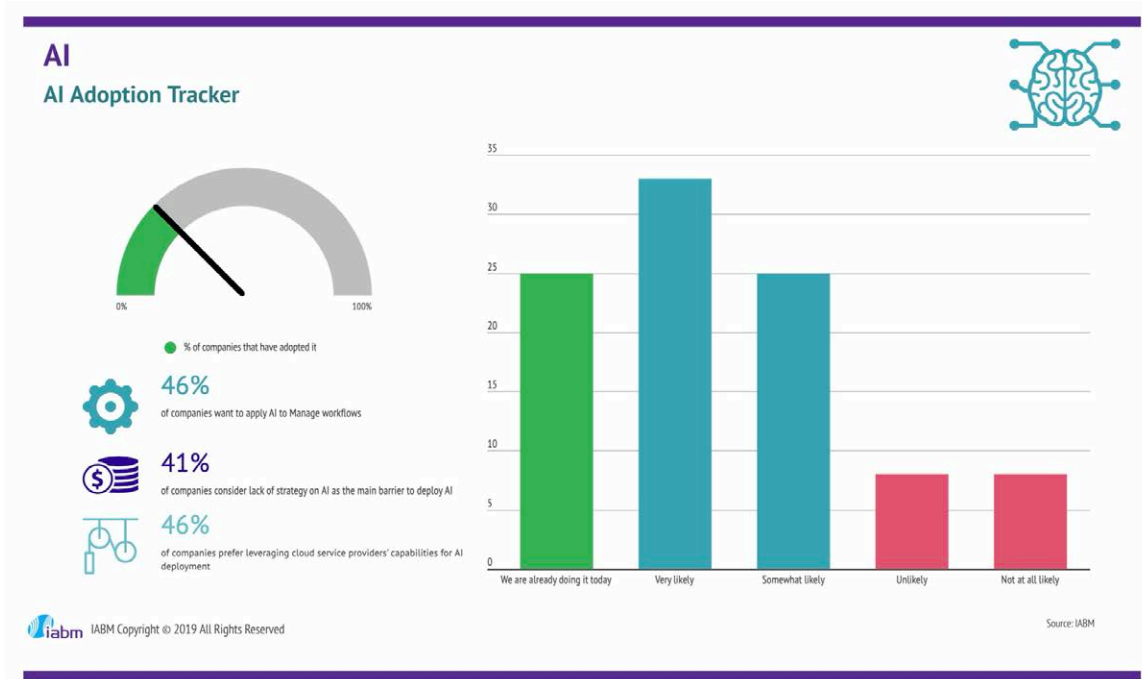
---

8) **Apply AI analysis that produces additional layers of TAM;**

- a. Text recognition
- b. Logo recognition
- c. Facial recognition
- d. Location recognition
- e. Sentiment analysis
- f. Scene detection
- g. Motion analysis

Any of these can be applied individually, or they can be applied conditionally based on business logic and metadata, or they can be concatenated, so that the product of an AI analysis/enrichment step can be used to determine what next step to take.

For instance: Not everyone can afford to apply AI analysis to their assets, even during a migration when the cost to do so is generally at its lowest. However, it may not be necessary to apply AI to everything. Consider the value of selectively applying AI to more limited sets of content based on available metadata, including content type, date and existing description. Then consider concatenating enrichment/selection



processes so that the most expensive AI services are only applied to the smallest and most targeted selection of content. You might start by applying a first level of AI Speech to Text enrichment to content of a specific type in a specific date range. A second level of text and logo recognition could be applied based on the presence or frequency of specific key words in the STT transcript. A third level of facial recognition could be applied based on the existence of additional key words. Finally, a fourth level of sentiment analysis could be applied based on the existence of interaction between two specific speakers. As these four AI processes are applied selectively, and only to a subset of the content produced by the process before them, not only does this result in highly focused output, but also reduced cost.

Technology moves quickly, in particular regards to AI and ML services. What might be the best service or fit with your business today might not be nine months from now. A Kumulate ISM gives you the ability to pick and choose the services that are most relevant and valuable to your business, right now – and the ability to make a different selection next week or next month. Two elements will allow you to assemble your customer workflow: 1) flexible business logic used to apply the chosen tools, and 2) the Kumulate ISM's relationships with a wide and diverse number of AI service providers.



## Conclusions and Takeaways

**Planning is everything**, and generally isn't something you'll be able to do entirely on your own. Seek advice from appropriate sources.

**There are myriad considerations...**

...around cost, content usage, storage tiering, added services (AI and ML), existing systems and many other aspects. Take the time to work these out.

**Migrations don't have to be painful.** Take the opportunity to use AI/ML to add value to your assets, create proxies or normalize your content.

Don't underestimate **the role of the ISM**, before, during and after the migration. The meshed view of all storage tiers, connected and an integral part of all workflows and operations, is key to **optimized content management**.

Migrations of large media repositories, and all of the associated metadata is a significant undertaking, and there's a lot to think about. But if you plan properly, take the time to evaluate existing resources, and make sure that you have the relevant tools and services as part of the project, migrating from LTO to cloud can be smooth, disruption-free, and a pathway to truly optimized content management.

---

Thanks for reading, we hope you found this document useful.

If you have any questions around migrations, content management, or workflow orchestration, simply drop us an email to [info@masstech.com](mailto:info@masstech.com) and we'll be happy to help. And you can find more information at [www.masstech.com](http://www.masstech.com)