ACCELERATING DIGITAL TRANSFORMATION BY MODERNIZING CONTENT SERVICES THROUGH CLOUD ADOPTION

Abstract

In today's digital-driven world, enormous amounts of content are being generated each day. Businesses on their digital transformation journeys are realizing the power of content in making business decisions. What was considered idle data in repositories is now becoming a powerful source to influence business direction. In this context, it is imperative for organizations to implement effective Content Services to unlock the full potential of organizational content.

This white paper establishes the need for Content Services modernization and highlights the advantages of adopting a cloud-based approach to accelerate enterprise digital transformation.

Infosys®
Navigate your next

Background

Businesses rely on multiple forms of information to make decisions such as whether to onboard a customer, approve a loan, and so on. In today's contactless world, digitized documents are becoming the primary source of information for decision-making on customer related business processes. By digitizing documents and automatically extracting information from the document, business decisions and the entire business process can be automated.

Consider a simple example. In a Know Your Customer (KYC) application, the Content Services capabilities required are optical character recognition (OCR), storage, and AI capabilities.

- Cloud service providers (CSPs) offer OCR as a service eliminating the need to buy a complete scanning solution for implementing OCR.

- Cloud storage is cheaper and offers tiered storage such as hot, cold, and archive to further optimize the storage cost.

- Cloud vendors also offer AI services for advanced content processing such as auto-classification and extraction.

All these services can be aggregated for workflow automation or pre-filling the metadata when a customer is uploading documents. This example illustrates how businesses can leverage cloud-based services offered by CSPs to improve the effectiveness of their Content Services.

Businesses can accelerate their digital journey by updating or adding new capabilities in the Content Services landscape to reap the benefits of automated decision-making. Investing in cloud-based services to enhance the power of their Content Services is a smart and cost-effective way to maximize the positive outcomes of digital transformation.
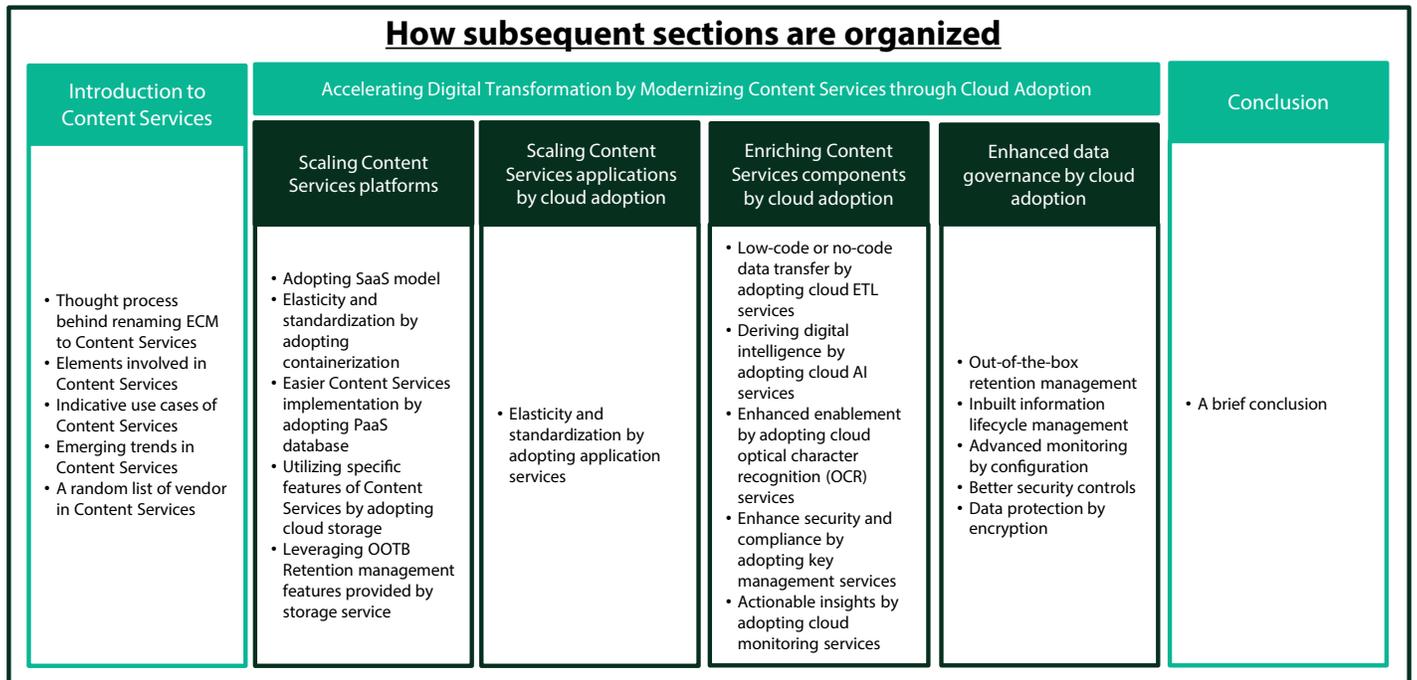
How subsequent sections are organized

Introduction to Content Services	Accelerating Digital Transformation by Modernizing Content Services through Cloud Adoption				Conclusion
	Scaling Content Services platforms	Scaling Content Services applications by cloud adoption	Enriching Content Services components by cloud adoption	Enhanced data governance by cloud adoption	
• Thought process behind renaming ECM to Content Services • Elements involved in Content Services • Indicative use cases of Content Services • Emerging trends in Content Services • A random list of vendor in Content Services	• Adopting SaaS model • Elasticity and standardization by adopting containerization • Easier Content Services implementation by adopting PaaS database • Utilizing specific features of Content Services by adopting cloud storage • Leveraging OOTB Retention management features provided by storage service	• Elasticity and standardization by adopting application services	• Low-code or no-code data transfer by adopting cloud ETL services • Deriving digital intelligence by adopting cloud AI services • Enhanced enablement by adopting cloud optical character recognition (OCR) services • Enhance security and compliance by adopting key management services • Actionable insights by adopting cloud monitoring services	• Out-of-the-box retention management • Inbuilt information lifecycle management • Advanced monitoring by configuration • Better security controls • Data protection by encryption	• A brief conclusion

Figure 1 – Organization of information in this paper

3 Introduction to Content Services

Content Services can be defined as any set of services or microservices that interacts with single or multiple content repositories to deliver one or more business use cases. Content can be of different types such as MS Word documents, and spreadsheets, or even audio/video files.

In early 2017, Enterprise Content Management (ECM) started being rebranded as Content Services. The primary reason for the change was the paradigm shift in how businesses wanted to use content, and ECM could no longer fulfill the dynamics around content-centric implementations that businesses wanted. Figure 2 summarizes the shift from ECM to Content Services.

ECM to Content Services - Modernization Path

Enterprise Content Management	Content Services
Focus on content management	Focus on content utilization
Manual process	AI-powered automatic workflow processing
Linear workflow	Collaborative and extended workflow
Single repository	Federated repositories
Monolithic architecture and self-contained repository	Modular or microservices-based architecture
Mostly on-premises	Cloud or hybrid deployments

Figure 2 – ECM vs Content Services

3.1 Elements of Content Services

Content Services comprise three major elements:

1. Content Services platforms represent the products or solutions deployed to create document repositories and provide fundamental features around content. This includes technologies that provide content capture, management, and delivery capabilities. For example, NewGen offers OmniDocs and OpenText offers Content Server as one of their products.

2. Content Services applications represent the business application, user interface, or other solutions which consume or operate on content. For example, a user interface connecting to a repository for uploading and retrieving content is a Content Services application.

3. Content Services components or enablers are solutions deployed to enrich the capabilities of Content Services. For example, any Extract-Transform-Load (ETL) service used to ingest documents into a repository can be considered a Content Services component.

Content Services

Content Services applications:
- Web interface
- Mobile interface
- External applications

APIs

Content Services platforms:
- Content Services products
- Metadata — Database
- Binary files — Storage

Content Services components:
- AI services
- OCR services — Enrichment enablers
- Key management — Security enablers
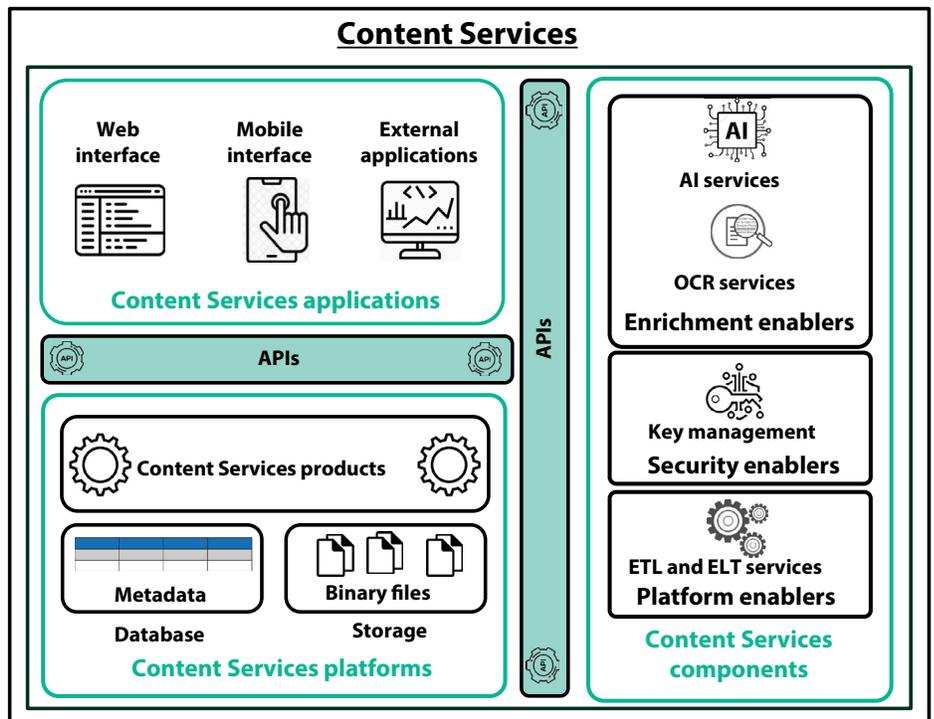- ETL and ELT services — Platform enablers

Figure 3 – Content Services ecosystem

Additional elements include an API layer which facilitates the interaction between the three Content Services elements. Figure 3 represents a Content Services ecosystem.

3.2 Content Services vendors

To implement Content Services solutions, several commercial as well as open-source products are available. Most vendors offer a complete Content Services suite as a set of products. Table 1 lists a few vendors and their Content Services offerings.

OpenText	OpenText is used by organizations across the world irrespective of the industry vertical. OpenText Extended ECM and OpenText Magellan are two products that are unique in their Content Services capabilities.
NewGen	NewGen's OmniDocs suite comprises modules that offer end-to-end implementation of Content Services. NewGen mainly focuses on formal business administration and operational use cases (customer-centric business solutions).
Hyland	Hyland with its flagship product, OnBase, offers a wide range of capabilities. Hyland acquired Alfresco and Nuxeo recently to strengthen their position in the Content Services space.
IBM	IBM has a wide range of products that offer end-to-end Content Services solutions.

Table 1 – Content Services solutions in the market

Content Services can be implemented using products from multiple vendors. For example, NewGen can capture content while OpenText serves as a content repository.

3.3 Content Services and unstructured content

Ninety percent of data managed today is unstructured and is growing at the rate of 55 to 60 percent per year. In other words, most of the content generated in any organization is unstructured. Digitizing and structuring the content is one of the key steps in digital transformation. Organizations on the path to digital transformation must deploy services and capabilities to enable effective use of content.

Unstructured content refers to information that does not have a pre-defined data model or is not organized in a pre-defined manner. Most common examples of unstructured data are images, MS Office documents, graphics and drawings, web pages, e-mail, videos, and rich media assets.

Structured content is content that is tagged and stored in an organized, searchable, and predictable way. Architecturally, structured content is saved in a database and unstructured content is stored on a file or network storage.

Content Services can manage both structured and unstructured content. While configuring Content Services platforms, metadata or a list of attributes are defined as a template. The objective of defining a metadata template is to make the content easily identifiable. Content Services products manage the link between unstructured content and its metadata. When unstructured content is ingested into Content Services, the platform stores the binary files and uses the storage path to create a link between the stored file and its metadata. The metadata and the link are stored in the database. When specific content needs to be retrieved, Content Services use the link to locate and display the corresponding piece of content.

3.4 Use cases for Content Services

Table 2 summarizes a few use cases for Content Services.

Industry vertical	Use cases	Content service capabilities	Benefits
Finance	Financial institutions can streamline the account opening process by implementing contactless and smooth onboarding experience for customers through end-to-end digitalization. Customers can upload documents digitally which can be verified automatically, thereby completing the end-to-end verification process quickly and with minimal manual intervention.	Capture, document management, BPM	Financial institutions can offer new generation digital experiences to their customer and stay ahead of the competition
Insurance	Insurance organizations can focus on process automation and decision-making by extracting information from documents to settle claims.	Capture, document management, BPM, extraction using AI	Automatic processing of workflow reducing manual processing significantly
Finance, Insurance	Customers need to be communicated with important events throughout their journey, such as changes in regulation, change in interest rate, etc. Personal greetings, such as birthday wishes also helps building goodwill.	Customer Communication Management	Effective customer communication helps in retaining a customer and helps in improving customer engagement.
Utilities	Content Services offers a solution to store product documentation and features such as versioning, lifecycle management, authorized access etc. By integrating Content Services with business application deployed on multiple sites, organizations can leverage single source of truth and eliminate redundant processes.	Document management, BPM	Productivity increases when field engineer can quickly retrieve latest version of product documentation
All verticals	Identify PI information from content and protect the PI information using proper record management policies	Capture, document management, BPM, extraction using AI	Regulatory compliance

Table 2 – Use cases for Content Services

3.5 Emerging trends in Content Services

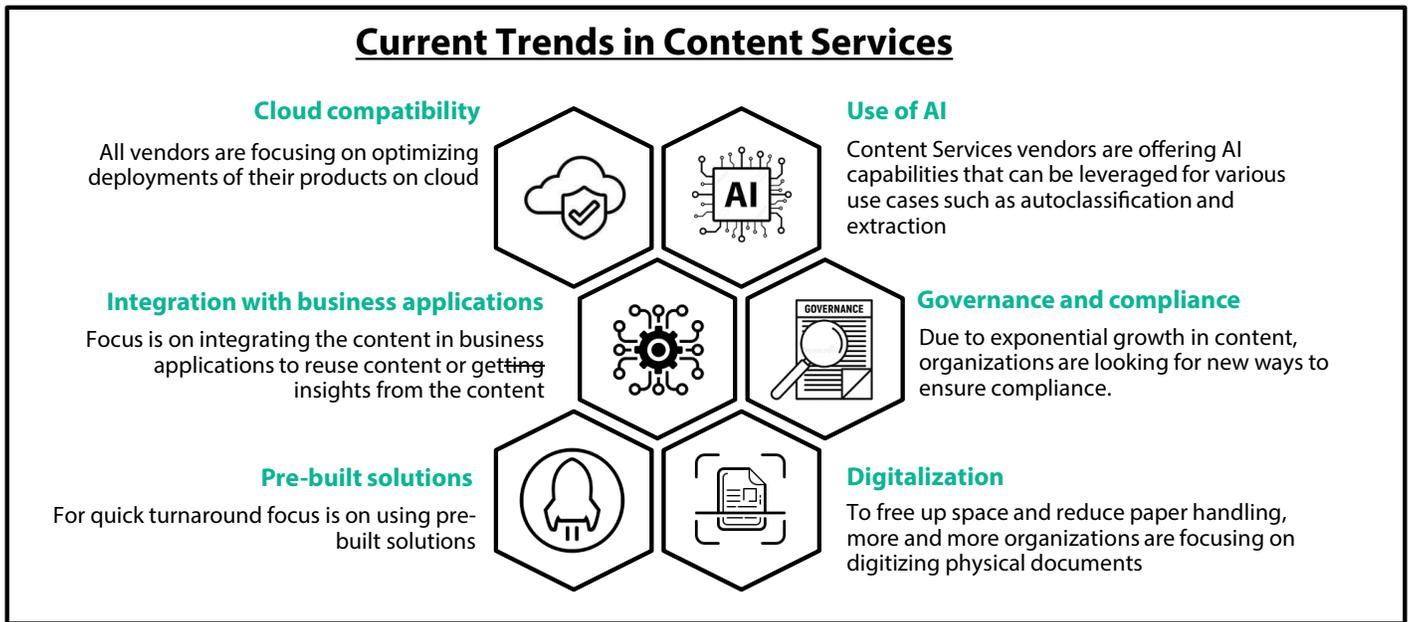Some emerging trends in Content Services are summarized in Figure 4.

Current Trends in Content Services

Cloud compatibility
All vendors are focusing on optimizing deployments of their products on cloud

Use of AI
Content Services vendors are offering AI capabilities that can be leveraged for various use cases such as autoclassification and extraction

Integration with business applications
Focus is on integrating the content in business applications to reuse content or getting insights from the content

Governance and compliance
Due to exponential growth in content, organizations are looking for new ways to ensure compliance.

Pre-built solutions
For quick turnaround focus is on using pre-built solutions

Digitalization
To free up space and reduce paper handling, more and more organizations are focusing on digitizing physical documents

Figure 4 – Emerging trends in Content Services

4 Modernization of Content Services Using Cloud Adoption

Organizations have accelerated their cloud adoption in the pandemic era. More and more businesses are adopting cloud-based digital solutions either progressively or fully.

This section covers some of the services provided by cloud technologies to modernize Content Services and accelerate digital transformation.

4.1 Scaling Content Services platforms

Cloud service providers (CSPs) offer various services that can be leveraged by Content Services platforms to scale and handle a high volume of content.

4.1.1 Adopting the SaaS model

Most prominent vendors offer content service capabilities based on the SaaS model. Businesses with lean requirements and without the required skillsets prefer to adopt Content Services as a service (CSaaS).

Advantages of CSaaS include low capex and maintenance costs, assured high availability, and business continuity. Adopting CSaaS can help small organizations accelerate their digital transformation and save on capital cost.

However, businesses must use features provided by the vendor and there is very little scope for customization. Security is one of the major concerns in adopting CSaaS models. In CSaaS, product or platform related issues are addressed quickly but there is a perception that debugging application-related issues is difficult.

Box and Dropbox offer Content Services exclusively on the cloud. Leading vendors such as OpenText, IBM, Nuxeo (now Hyland), and NewGen provide CSaaS as well as on-premise deployment options. For implementing advanced use cases such as process automation and text analytics, organizations must depend on services providers.

4.1.2 Elasticity and standardization by adopting Containerization

Containerization and orchestration are standard services provided by most CSPs for Content Services deployments.

While using Content Services, capacity planning must be done for peak load and for various workload models and transactional mix. Despite planning, resources remain idle during non-peak scenarios and need to be paid for.

Key benefits:

- Cloud deployments provide a cost advantage and open the possibilities of autoscaling and serverless architectural options.

- Leveraging containerization leads to increased portability, guaranteed high availability, consistent deployment, and easier CI/CD integration.

- Simplified maintenance through consistent responses and easy upgrades reduces overall maintenance cost and enables compliance to infrastructure-related SLAs.

- Using containerized Content Services in Kubernetes cluster enables the implementation of on-premises deployments.

Vendors such as OpenText, NewGen, and Nuxeo offer their core content server product as an image which can be containerized.

Through containerization of Content Services platforms, organizations can quickly set up their Content Services implementation without spending time on capacity planning or fulfilling high availability requirements. Content Services deployments can be easily integrated with the DevOps pipeline and other systems thereby accelerating digital transformation

If the Content Services product does not support containerization, autoscaling can be used as a workaround. Autoscaling is a service provided by CSPs using which an application can be deployed on multiple virtual machines (VMs) behind a load balancer and scaled according to preconfigured criteria. It provides similar elasticity but at the VM level. Use of containers and autoscaling simplifies lifecycle management of the product being used..

4.1.3 Easier content service implementation by adopting a PaaS database

Content Services platforms usually need an RDBMS database to store metadata. The performance of Content Services platforms depends on the performance of the backend database. Hence, from a performance and scalability point of view, using the right database is very important.

CSPs offer database services as Platform as a Service (PaaS). Availing PaaS services means one less component to manage, upgrade, and maintain. Coupled with the features of autoscaling or elasticity, the performance and scalability of the Content Services platforms are assured. Using a PaaS database helps in setting up Content Services quickly.

CSPs also offer document-based NoSQL databases as PaaS. Content Services platforms can harness the power of NoSQL databases to store and handle high volumes of data. Content Services vendors such as Nuxeo (now Hyland) support the use of NoSQL as the backend database.
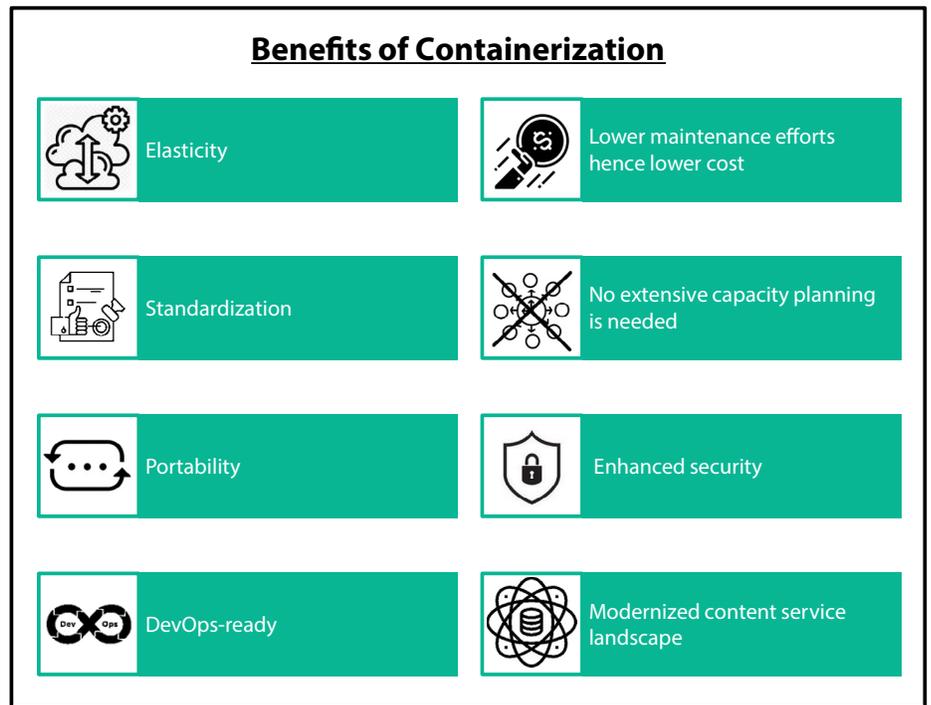

Figure 5 – Benefits of containerization

4.1.4 Utilizing Content Services features by adopting cloud storage

Content Services store binary files of content in file systems. CSPs offer cost optimized and highly scalable storage built on object store specifications. Content service vendors can leverage this storage to optimize cost, implement life cycle management, and utilize retention management features. Figure 6 illustrates the features of cloud storage.
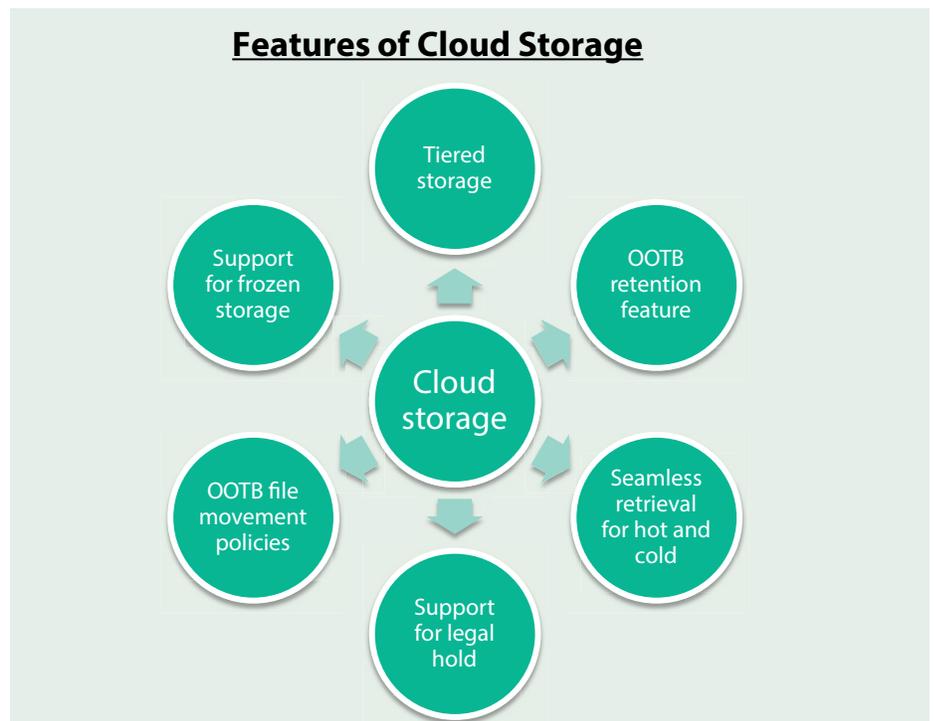

Figure 6 – Features of cloud storage

All cloud storage vendors offer out-of-the-box lifecycle management or tiered storage such as hot tier, cold tier, and archived tier.

- When a document is ingested into the Content Services platform, the binary file is stored in the hot tier.

- As content grows older, the file is accessed less frequently, it can be moved seamlessly to the cold tier by configuring OOTB storage policies. The file retrieval process remains unchanged. For instance, in Microsoft

Azure, moving old files to the cold tier can yield storage cost savings of up to 40%. There could be a performance tradeoff in content retrieval from the cold tier, but it will still provide real-time access.

- Old and unused content can be moved to the archived tier but may need to be retained for long durations for compliance reasons. Retrieval of content from the archived tier is not possible in real-time. For long term archival strategies, it is best to convert the content to PDF/A format.

Content Services can leverage these storage features to save on storage cost. In addition, Content Services platforms can leverage the object store feature to store extra attributes. Using this feature, when the documents are archived, only important attributes are stored in the database while all other attributes are moved to the object store for archival. This saves database space and resources.

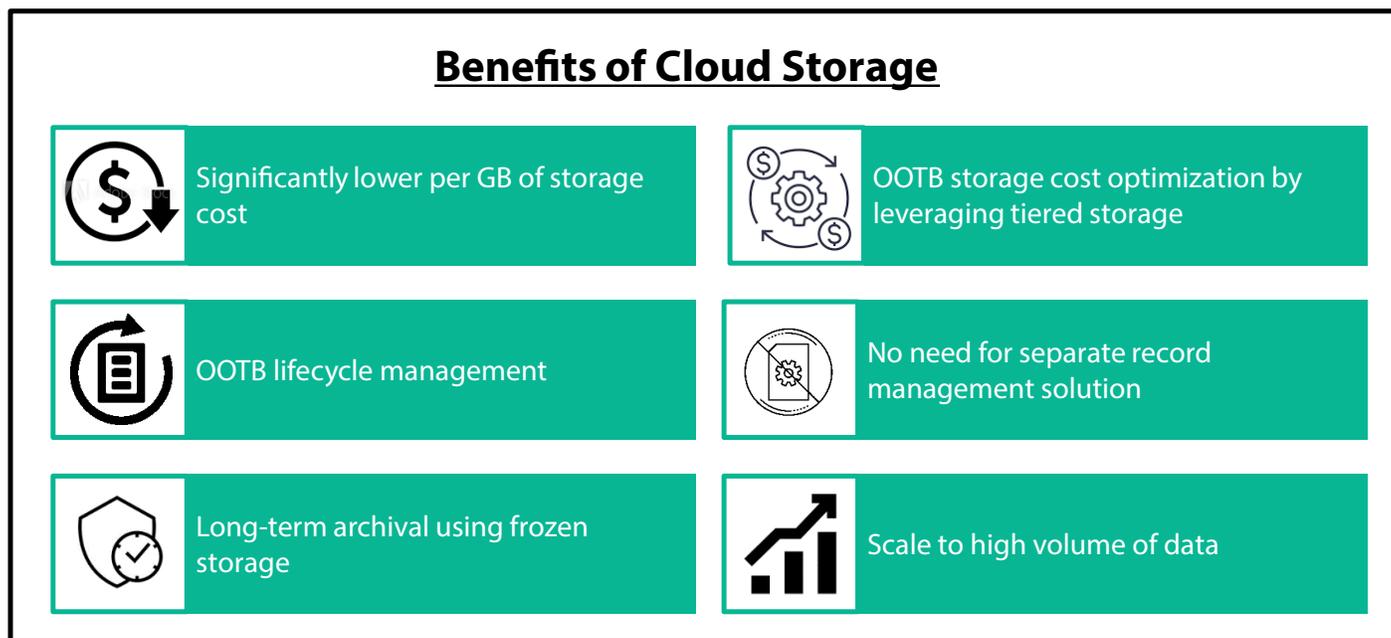Figure 7 lists the benefits of cloud storage.

Benefits of Cloud Storage

- Significantly lower per GB of storage cost
- OOTB storage cost optimization by leveraging tiered storage
- OOTB lifecycle management
- No need for separate record management solution
- Long-term archival using frozen storage
- Scale to high volume of data

Figure 7 – Benefits of cloud storage

4.1.5 Leveraging OOTB retention management features provided by cloud storage

Retention management is a feature advocated by records management through which records or binary files under retention are guaranteed to remain in the system, unaltered, for a specified time duration. Any request to modify metadata or update binary files of the records under retention are rejected.

Records under retention are also protected against accidental or intentional deletion.

Metadata stored in the database is always accessed through the database engine that authorizes all requests to operate on the metadata. Hence, metadata is protected, and retention is guaranteed with proper permissions.

It is difficult to guarantee the retention of documents. But CSPs such as AWS and Microsoft Azure provide out-of-the-

box retention management features at storage level. Once configured, binary files under retention become immutable. Any request to update or delete the binary file is rejected. Retention management policies can be configured with just a few clicks. Microsoft Azure also offers legal hold features under which documents needed for an ongoing investigation can be retained. This tailored records management feature helps achieve compliance at much lower costs.

4.2 Scaling Content Services applications

Content Services applications are the front-end used by business users to upload, retrieve, and modify content on Content Services platforms.

4.2.1 Elasticity and standardization by adopting application services

On-premises hosting of Content Services applications requires significant investments in terms of resources and infrastructure. Cloud vendors offer deployment of applications using application services. For example, Microsoft Azure offers Azure App Service and AWS offers AWS Elastic Beanstalk. Azure App Service is a fully managed service that offers features such as elasticity, quick time to market, CI/CD integration, less maintenance, security, and regulatory compliance at a cost which is up to 54% percent lower than the same application being deployed on-premises (https://azure.microsoft.com/en-in/services/app-service/#features).

4.3 Enriching Content Services components

This section discusses the modernization of Content Services components using cloud adoption.

4.3.1 Low-code or no-code data transfer using cloud ETL Services

Extract-Transform-Load (ETL) is used to ingest content into a Content Services platform. CSPs provides ETL services that can be leveraged to set up low-code or no-code ETL pipelines. AWS provides AWS Glue or AWS pipeline whereas Microsoft Azure provides Azure Data Factory for setting up an ETL pipeline. Using these ETL services, a template ETL pipeline or master pipeline can be set up which can be tweaked for setting up new pipelines. ETL services are available as serverless components and can scale automatically based on the load. Also, features such as monitoring dashboards, schedulers, exception handlers, and re-run capabilities are available out of the box. Since no infrastructure maintenance is involved, overall cost is low, and the Content Services landscape is simplified.

ETL services offered by cloud vendors can accelerate digital migration to new platforms and solutions by providing a no-code or low-code framework to migrate the data.

Figure 8 lists the benefits of adopting cloud ETL services.

Benefits of Adopting Cloud ETL Service

- ETL standardization
- Additional infra is not needed
- Low-code/no-code set-up
- Custom tasks for specific requirement
- Quick time to market
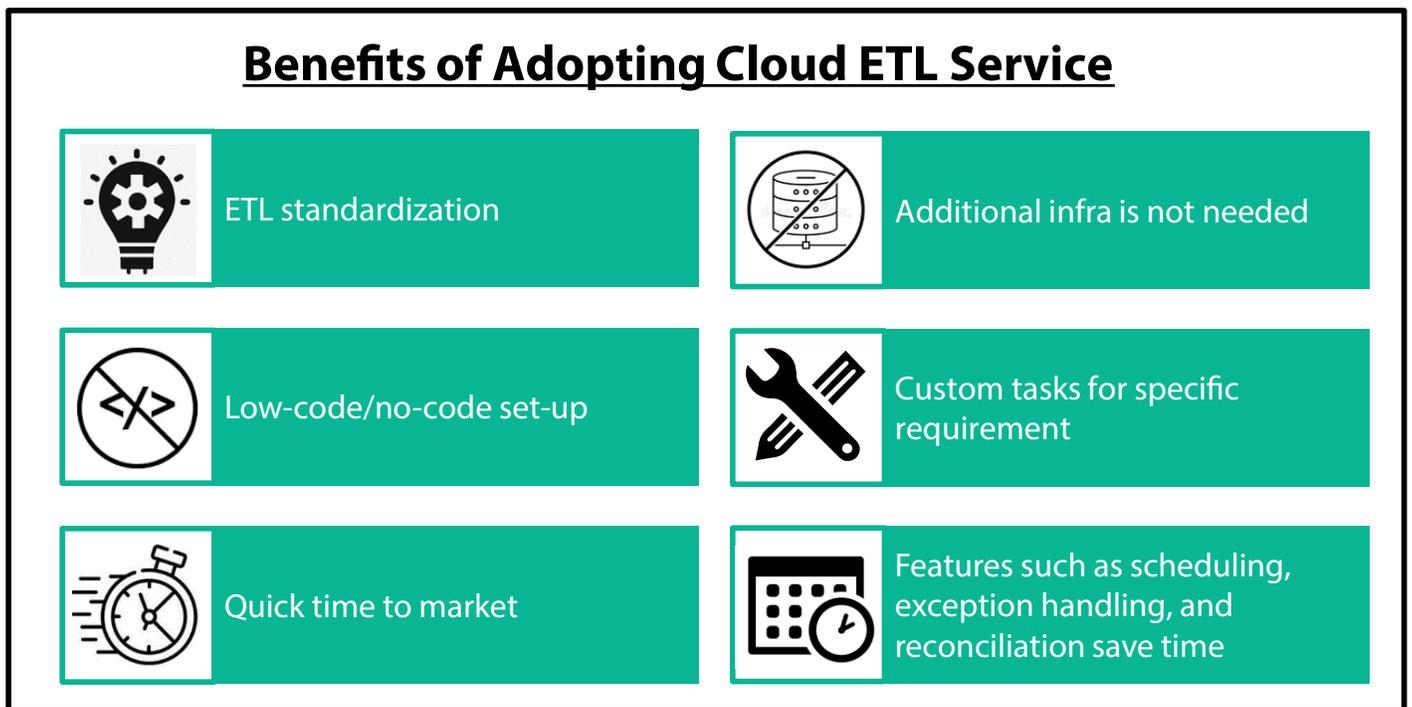- Features such as scheduling, exception handling, and reconciliation save time

Figure 8 – Benefits of Cloud ETL services

4.3.2 Deriving digital intelligence by adopting cloud AI services

AI is becoming an integral part of Content Services and CSPs are facilitating proliferation of AI use cases by providing feature-rich AI services.

Using AI, an organization can offer an enhanced experience to customers as well as employees. This includes pre-filling contextual metadata values when a document is being uploaded and creating intelligent summarization of large documents. Traditionally, Content Services repositories have been perceived as idle repositories with no valuable information provided to business. Using AI, critical information hidden in documents can be unlocked to provide businesses with cognitive capabilities.

Since Content Services deal with documents, natural language processing (NLP) related use cases such as entity extraction, auto-tagging, and auto-classification are more relevant. Chatbot is the most popular AI solution in the industry powered by NLP.

Figure 9 lists a few use cases for NLP in Content Services.

NLP Use Cases in Content Services

Auto verification

Document verification intelligence is provided by classification

Digitized forms autofill

Pre-filling intelligence is provided by metadata extraction

Summarization

Summarization intelligence is provided by extracting important terms and sentences/paragraphs

Digital KYC

Document identification and verification intelligence is provided by classification and extraction of metadata

Workflow automation

Workflow automation is enabled by extracting metadata from documents

Improving discoverability

Document findability is enhanced by extracting important terms and tagging these back to documents
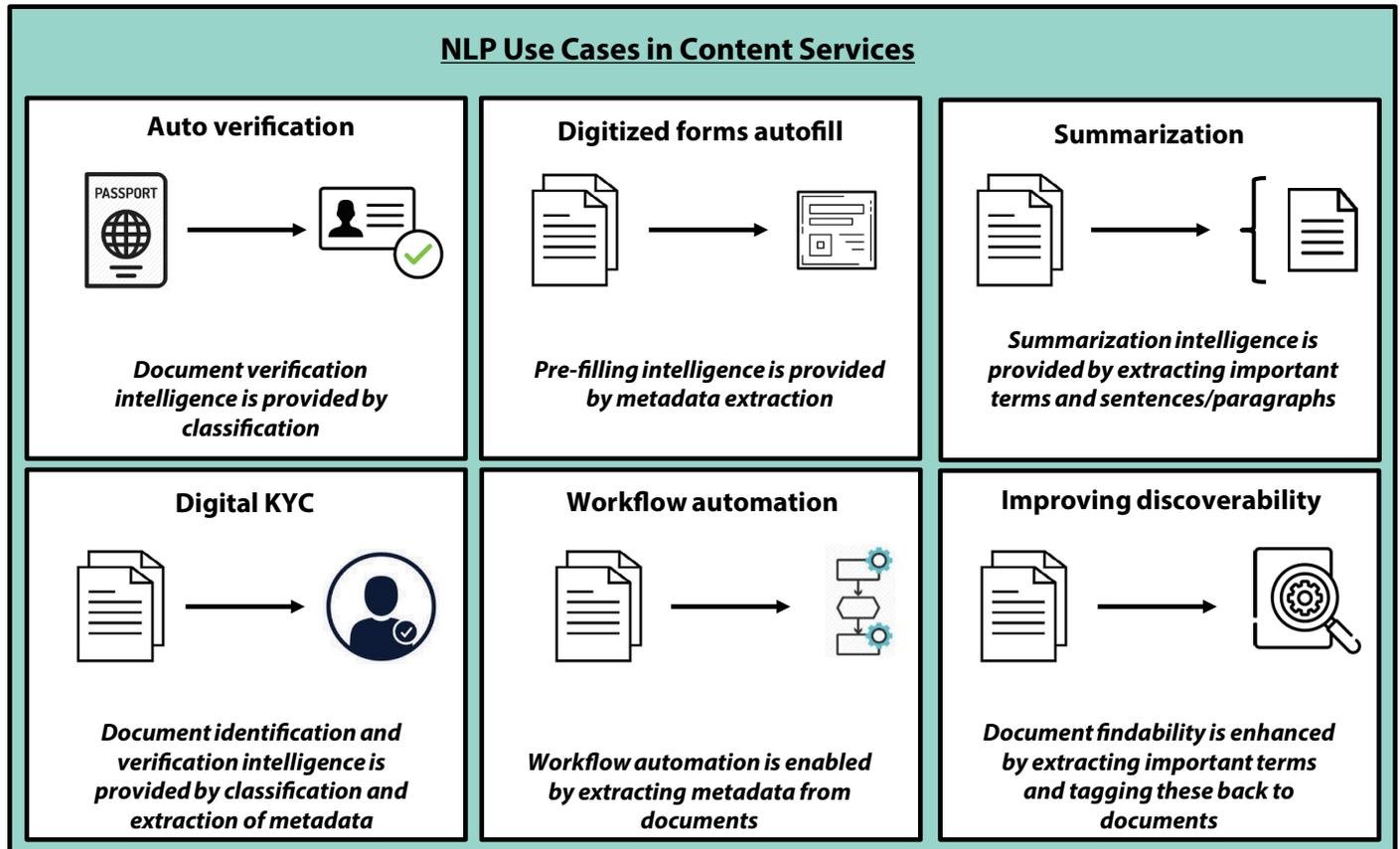
Figure 9 – Use cases for NLP in Content Services

For standard use cases such as sentiment analysis, topic detection, and language detection, OOTB services can be used. For advanced use cases such as organization specific entity extraction and classification, the NLP engine needs to be trained to a level of acceptable accuracy before being deployed.

The key benefits of using cloud-based NLP solutions are no-code or low-code set-up, quick time to market, and low cost of deployment. These benefits help organizations offer digital use cases to their customer quickly compared to traditional Content Services.

4.3.3 Enhanced enablement by adopting cloud OCR services

Optical character recognition or OCR is a capability to extract text from images. Earlier, OCR was available only through scanning software. CSPs are now offering dedicated OCR services. Intelligent content recognition (ICR) is a variation of OCR that deals with the recognition of handwritten text.

With the advent of digitization, usage of digital content has grown significantly. Digital content with images must be processed by OCR software to extract text. The extracted text can be processed by an NLP engine to intelligently extract entities or auto classify the documents.

Using cloud-based OCR and ICR services, text from files can be extracted and processed further depending on use cases. OCR and ICR services can be used even if the application is running on premise. Using OCR and ICR services coupled with cloud-based AI services, use cases such as entity extraction, auto-tagging, and auto-classification can be implemented quickly.

4.3.4 Improved security and compliance by adopting key management services

In most Content Services applications, passwords are encrypted and stored in configuration files. CSPs provide key management services to deal with such encrypted data. A key management service can store passwords and certificates, and only authorized users or applications can retrieve the information. An application can also be authorized to access the key. This enables an application to register itself to access the keys and use it. For security reason, key management services automatically generate the audit trail..

Some CSPs provide advanced features such as primary key and secondary keys, re-cycling keys, and automatic syncing of new keys with the key manager. Primary keys

are set to recycle at pre-defined intervals. Till the new recycled key gets reflected into the key manager, applications can use the secondary keys. Once the new primary key is available in the key manager, the application can start using the primary key.

4.3.5 Actionable insights by adopting cloud monitoring services

Typically, application monitoring involves selection, procurement, and installation of the required tool, configuration, and deployment of associated agents.

However, CSPs offer monitoring services in a much simpler way. With a few configurations, monitoring services can start running and monitor required applications and services. Monitoring agents can be installed on VMs if log analytics are needed for advanced monitoring. Integration with AI techniques are also available for advanced use cases such as detecting anomalies and predictive maintenance.

4.4 Enhanced data governance with cloud adoption

Content Services can implement the following data governance strategies easily by adopting cloud services:

- Out-of-the-box retention management – Cloud storage offers retention management OOTB. Content Services

can leverage this feature to offer tailored records management and implement data governance reduced effort and at much lower cost.

- Inbuilt information lifecycle management – In cloud storage, binary files can be moved from hot tier to cold tier by configuring data movement policies. Using this feature, older binary files can be automatically moved to cold tier. This saves storage cost by almost 50%. Cloud storage also provides archived tier which can hold data for long durations at very low cost.

- Advanced monitoring – There is no governance without monitoring. Using cloud services, monitoring is easy and relevant information can be generated and presented on a dashboard. IT staff and business users can take corrective action as needed.

- Better security controls – In cloud deployments, an application's access to another application can be easily controlled. This is achieved using managed identities in Microsoft Azure and IAM roles in AWS. Using role-based access control (RBAC), granular level access control can be implemented.

- Data protection by encryption – Data is always encrypted by the CSP's key. Businesses can choose to encrypt the data with additional keys.
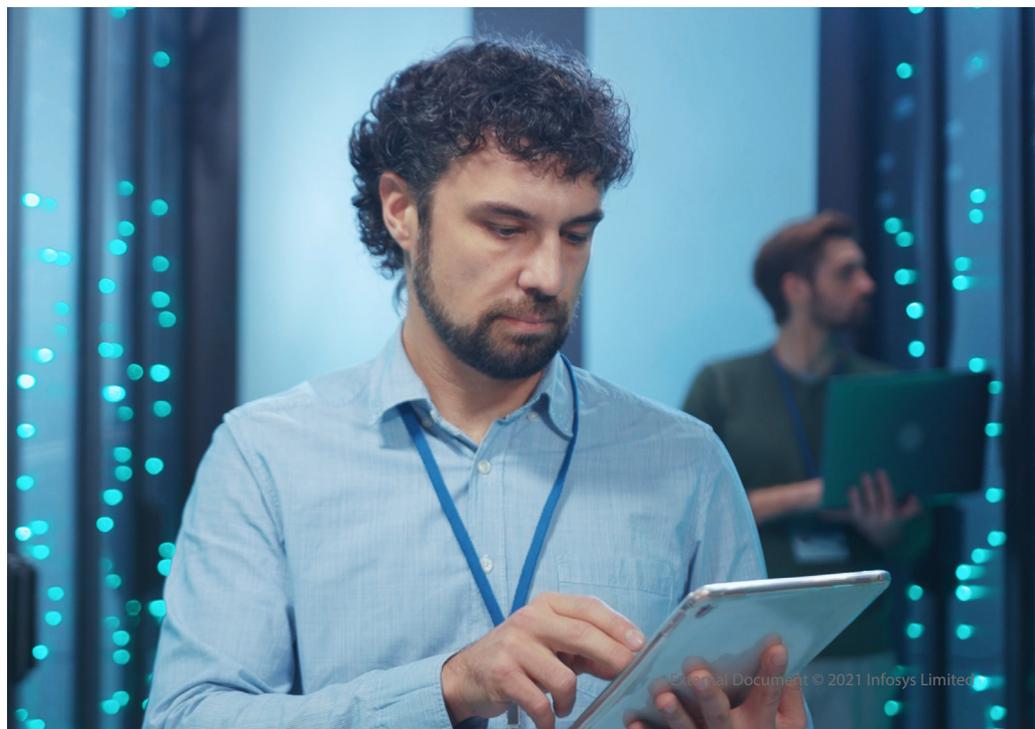
Conclusion

While adopting cloud strategies for Content Services, a lift and shift approach that involves setting up VMs, installing Content Services products, and hosting applications might not yield the aspired level of cost benefits or business values. CSPs offer a range of cloud-based solutions that can be used by Content Services for simplifying and modernizing the landscape.

Cloud services can also be leveraged for a select set of capabilities. For example, core Content Services can continue on-premises and CSPs can provide capabilities such as AI, OCR, and text analytics. The real value is in leveraging appropriate PaaS and SaaS offerings from CSPs based on the Content Services landscape and cloud-native aspirations of the enterprise.

Use cases needed for digital transformation such as process automation, content analytics, and low-code/no-code solutions can be implemented much faster in a cloud than in a traditional landscape. Therefore, cloud adoption not only modernizes Content Services implementation but also accelerates the digital transformation of the organization.
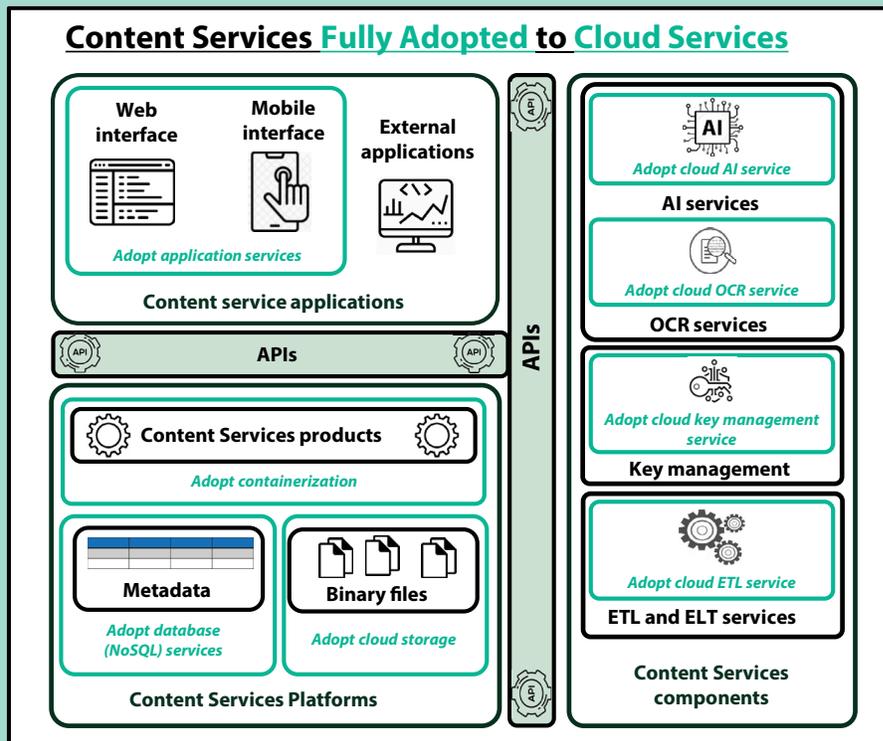
Content Services Fully Adopted to Cloud Services

Content service applications
- Web interface
- Mobile interface
- *Adopt application services*
- External applications

APIs

Content Services Platforms
- Content Services products
- *Adopt containerization*
- Metadata — *Adopt database (NoSQL) services*
- Binary files — *Adopt cloud storage*

APIs

Content Services components
- AI services — *Adopt cloud AI service*
- OCR services — *Adopt cloud OCR service*
- Key management — *Adopt cloud key management service*
- ETL and ELT services — *Adopt cloud ETL service*

Figure 10 – Serverless Content Services with cloud services adoption

If cloud services are used to their full potential, a modernized "serverless Content Services" architecture would appear as depicted in Figure 10.

Modernization by adopting cloud services is applicable to other enterprise tools and technologies and will also help accelerate digital transformation.

About the Author

Girish Pande

A senior technology architect with the Digital Marketing unit at Infosys

Has around 21 years of experience in information technology. He has played key roles in architecting and implementing end-to-end Content Services, moving Content Services to the cloud, refactoring and rearchitecting Content Services in the cloud, implementing NLP use cases, building automation solutions using RPA and BPM, and implementing enterprise search for various clients across the globe. Girish has an M. Tech. in Industrial Engineering & Operations Research from IIT Bombay.

He can be reached at girish_pande@Infosys.com.

References:

1. https://blogs.opentext.com/an-approach-towards-application-containerization/
2. https://documentmedia.com/article-2958-The-Disruption-of-Content-Services-The-Vision-for-2020-and-Beyond.html

Infosys®
Navigate your next

For more information, contact askus@infosys.com