TECHNICAL WHITE PAPER
November 2024

VMware Cloud Foundation with Microsoft Azure Video Indexer Enabled by Arc

Table of contents

Note: This solution provides general solution design and deployment guidelines for running the Microsoft Azure Video Indexer enabled by Arc on VMware Cloud Foundation. The hardware used in this paper is not strictly tied to the solution architecture. The solution architecture applies to any compatible hardware platforms running VMware Cloud Foundation.

Executive Summary

Business Case

The integration of VMware Cloud Foundation with Microsoft Azure Video Indexer, facilitated by Azure Arc, provides organizations with a powerful solution for running Azure workloads securely on-premises. This approach allows businesses to leverage the advanced capabilities of Azure Video Indexer while maintaining control over their data in a private environment. By utilizing VMware's Private AI framework, organizations can implement AI-driven analytics on their video content without compromising security or compliance, effectively transforming their multimedia assets into actionable insights.

With VMware Cloud Foundation, enterprises can create a flexible and scalable infrastructure that supports various workloads, including AI applications. This integration enables organizations to deploy GPU-accelerated and non-GPU-accelerated environments optimized for all kinds of AI tasks, ensuring that they can harness the full potential of their data within a secure framework. The combination of VMware's robust management tools and Azure's extensive cloud services fosters an environment where businesses can innovate rapidly while adhering to stringent data privacy standards.

Moreover, this solution enhances operational efficiency by allowing organizations to utilize familiar VMware and Azure tools for managing both on-premises and cloud-based workloads. This continuity minimizes the need for extensive retraining of IT staff and accelerates the adoption of new technologies. As organizations increasingly seek to modernize their IT infrastructure, the ability to run Azure workloads securely on-premises through VMware Cloud Foundation becomes a critical differentiator in maintaining competitive advantage.

In summary, the collaboration between VMware and Microsoft empowers businesses to maximize the value of their video content while ensuring that sensitive data remains protected. By integrating Azure Video Indexer within a secure VMware environment, organizations can drive innovation and enhance productivity, ultimately leading to improved business outcomes and customer engagement.

Audience

This VMware Validated Solution paper is intended for the following audiences:

- Enterprises and users who are using Azure cloud who also have a business or technical need to run AI services on-premises.
- Corporate CTOs and CIOs who are interested in learning more about how Broadcom and Microsoft can work together to bring Azure AI services on-premises.
- vSphere VI administrators who are familiar with VMware virtualized infrastructure and are looking to deploy and manage AI workloads in a private virtualized environment.
- Any other engineer/operator/end-user who are interested in AI workloads, Kubernetes, VMware Cloud Foundation and have a good understanding about VMware Cloud Foundation, vSAN, NSX, Cloud Native Storage (CNS), Container Storage Interface (CSI), Kubernetes, and AI.

Technology Overview

VMware Cloud Foundation

VMware Cloud Foundation is an integrated software stack that combines compute virtualization (VMware vSphere), storage virtualization (VMware vSAN), network virtualization (VMware NSX), and cloud management and monitoring (VMware Cloud Foundation Automation and VMware Cloud Foundation Operations, formerly known as VMware Aria Automation and VMware Aria Operations) into a single platform that can be deployed on-premises as a private cloud or run as a service within a public cloud. This documentation focuses on the private cloud use case. VMware Cloud Foundation bridges the traditional administrative silos in data centers, merging compute, storage, network provisioning, and cloud management to facilitate end-to-end support for application deployment. See Getting Started with VMware Cloud Foundation for details.

Microsoft Azure

Microsoft Azure is a comprehensive cloud computing platform offering over 200 services across categories such as compute, storage, networking, databases, AI, and IoT. It provides Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) solutions, enabling organizations to build, deploy, and manage applications and services through Microsoft-managed data centers. Azure's global infrastructure supports scalable deployment options with a pay-as-you-go pricing model, making it suitable for businesses of all sizes.

Microsoft Azure Arc

Microsoft Azure Arc extends Azure's capabilities beyond the cloud, enabling organizations to run Azure services on-premises, in other clouds, or at the edge—from a single interface. With Azure Arc, users can deploy Azure services on any infrastructure, including Kubernetes clusters on VMware Cloud Foundation, while leveraging GitOps principles for configuration management.

Microsoft Azure Video Indexer

Microsoft Azure Video Indexer is an AI-driven service that extracts valuable insights from video and audio content. By automatically generating metadata through visual and audio analysis—including face detection, speech-to-text transcription, and sentiment analysis—Video Indexer enhances the discoverability and accessibility of multimedia assets. This service is particularly beneficial for industries such as media and entertainment, education, and enterprise communications.

The platform offers a user-friendly web interface along with a REST API for seamless integration into existing workflows. By leveraging Azure's robust infrastructure, Video Indexer can efficiently process large volumes of video content at scale, empowering organizations to create engaging user experiences and streamline content management.

Microsoft Azure Video Indexer enabled by Arc

Microsoft Video Indexer Enabled by Arc combines the advanced capabilities of Azure Video Indexer with the flexibility of Azure Arc. This integration allows organizations to deploy video indexing solutions on-premises or at the edge while maintaining consistent management through Azure's unified interface. It addresses critical needs such as data sovereignty and low-latency processing by enabling local analysis of sensitive video content.

This hybrid approach is ideal for various sectors, including media production, healthcare, manufacturing, and public services. By leveraging both cloud-based and on-premises resources, organizations can optimize workflows while ensuring compliance with data privacy regulations. This solution represents a significant advancement in making AI-powered video analysis accessible across diverse IT environments.

Solution Configuration

This section introduces the following resources and configurations:

Well-architected Solution

- Hardware resources
- Software resources
- VMware Cloud Foundation Deployment
  - vSAN File Service Deployment
  - Tanzu Kubernetes Grid Service (TKGs) Deployment
  - Tanzu Kubernetes Cluster Preparation
- Azure
  - Connect on-prem Kubernetes Cluster via Azure Arc
  - Create a Microsoft Azure Video Indexer Account
  - Deploy Microsoft Azure Video Indexer Extension
- Access Video Indexer, Upload Videos and Get Insights

Solution Architecture

The VMware Cloud Foundation validation environment is composed of a standard architecture, including a management domain and one workload domain.

Different workload domains can serve different business purposes. The one-to-many mapping simplifies the overall management of the whole VMware Cloud Foundation environment.

The Azure Video Indexer solution will be deployed in the workload domain on top of TKGs Cluster. The TKGs cluster will connect to Azure Arc. Once connected to Arc, the Azure VI ARC Extension will deploy the Video Indexer solution on-prem.

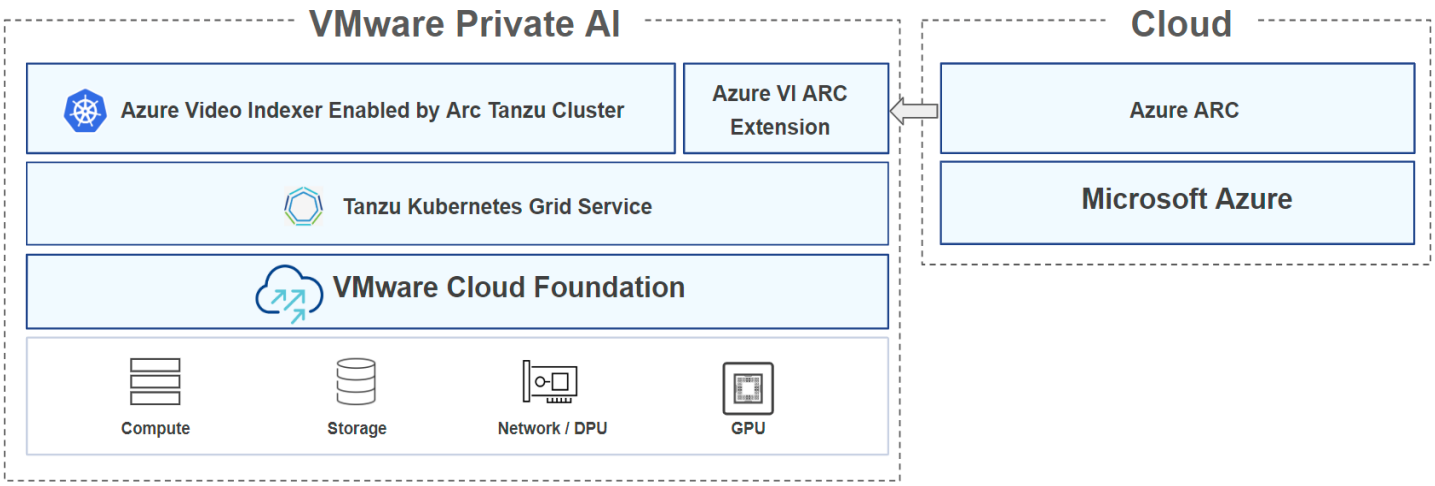


Figure 1: Architecture Diagram - VMware Private AI and Azure Arc

VMware Cloud Foundation Standard Architecture

For the standard architecture, we created a 4-node ESXi cluster for the VMware Cloud Foundation management domain, running management virtual machines and appliances.

Table 1. Management Domain VMs

VM Role	vCPU	Memory (GB)	VM Count
Management Domain vCenter Server	4	16	1
SDDC Manager	4	16	1
Management Domain NSX Manager	6	24	3
Workload Domain NSX Manager	12	48	3
Workload Domain vCenter Server	8	28	1

For the workload domain, we created another 4-node ESXi cluster with a separate NSX Fabric, deployed an NSX Edge Cluster, deployed vSAN File Service, and TKGs. To learn more about VMware Cloud Foundation domain architecture, click here.

Table 2. Workload Domain VMs

VM Role	vCPU	Memory (GB)	Storage	VM Count
Control Plane TKC	4	16	25	1
Worker Node TKC	48	256	100	2

Hardware Resources

For the standard architecture, we created a 4-node ESXi cluster for the VMware Cloud Foundation management domain, running management virtual machines and appliances.

In this solution, for the VMware Cloud Foundation workload domain cluster, we used a total of four Dell R740xd vSAN Ready Nodes. Please consult our Compute Design for VMware Cloud Foundation for more information.

Each ESXi node in the cluster had the following configuration, as shown in Table 3.

Table 3. Hardware Configuration used for the Workload Cluster

Property	Specifications
Server model name	Dell PowerEdge R740xd vSAN Ready Node
CPU	2 x Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz 24 Cores
RAM	384GB
Network Adapter	Dell HBA330

Storage Adapter	8
Disks	2 x Toshiba KPM5XMUG400G 400GB SSD as vSAN Cache Device 2 x WDC WUSTR1519ASS200 1.92TB SSDs as vSAN Capacity Devices

Software Resources

Table 4 shows the software resources used in this solution.

Table 4. Software Resources

Software	Version	Purpose
VMware Cloud Foundation	5.2	A unified SDDC platform that brings together VMware vSphere, vSAN, NSX, and optionally, VMware Cloud Foundation Automation, into a natively integrated stack to deliver enterprise-ready cloud infrastructure for the private and public cloud. See VMware Cloud Foundation 5.2 Release Notes for details.
Tanzu Kubernetes releases (TKr)	v1.28.8---vmware.1-fips.1-tkg.2	Microsoft Video Indexer enabled by Arc requires Kubernetes release 1.26 or higher running Ubuntu 22.04.
AzureCLIr	2.48.0	AzureCLI minimum required version to interact with Azure services.
nginx	1.11.2	Ingress Controller.

Table 5 shows the configuration used to create the Tanzu Kubernetes Cluster

Table 5. Tanzu Configuration

VMClass Name	Configuration	Purpose
guaranteed-large (Default Class)	4 x vCPUs, 16GB RAM	• Control Plane node (1 node). Can add 3 nodes for redundancy if required.
azure-vi-large (Custom Class)	48 x vCPUs, 256GB RAM, 100GB Storage	• Recommended requirement (2 worker nodes) • Ideal for larger deployments. • Process 10 videos in Basic Video Preset. • Custom VM classes must be provisioned prior to TKC deployment. Consult documentation for more information

Note: *No GPUs were used in this environment.*

Solution Deployment

VMware Cloud Foundation Installation

Follow the Official Document for detailed information about VMware Cloud Foundation installation steps. The key steps for VMware Cloud Foundation installation are as follows:

1. Deploy SDDC Manager and management workload domain through Cloud Builder.
2. Add ESXi hosts to SDDC.
3. Create a workload domain with the idle ESXi hosts.
4. Deploy TKGs on the workload domain cluster.

Deploying VMware vSAN File Service

To deploy vSAN File Service, we will enable the service on the workload domain cluster, and configure the necessary networking and domain settings. Enabling vSAN File Service allows Tanzu Kubernetes Cluster to use ReadWriteMany (RWX) storage class, a requirement for Video Indexer.

1. Enable File Service:
    a. Navigate to the vSAN cluster in the vSphere Client.
    b. Click on Configure > vSAN > Services.
    c. On the File Service row, click Enable.
2. Download the OVF File: Choose to automatically download the OVF file needed for deployment.
3. Configure Domain Settings:
    a. Specify a unique File Service domain name.
    b. Enter DNS server details (A and PTR records), subnet mask, and gateway information.
    c. Assign IP Addresses: Provide a pool of static IP addresses for the file servers, ensuring that the number of IPs matches the number of hosts in your cluster.
4. Finalize Configuration: Review all settings and click Finish to complete the setup.

It is also required to activate file volume support on the Supervisor Cluster.

1. In the vSphere Client, navigate to **Workload Management**.
2. Click the **Supervisors** tab and select the Supervisor to edit from the list.
3. Click the **Configure** tab and click **Storage**.
4. Select File Volume.
5. Activate file volume support.

Deploying VMware Tanzu Kubernetes Cluster

The first step in our integration is to deploy a Tanzu Kubernetes Cluster (TKC) on VMware Cloud Foundation. This cluster will serve as the foundation for our on-premises workloads and enable seamless connectivity with Azure services. The steps to deploy a Tanzu Kubernetes Cluster are:

1. Create Namespace within vSphere.
    a. Log into vSphere Client and navigate to the Workload Management section.
    b. Select the Supervisor cluster where you want to create the namespace.
    c. Launch the Create Namespace wizard:
        i. Choose a name for the namespace (must be DNS compliant).

ii. Select the workload network.

2. Configure namespace settings:
   a. Permissions
      i. Add users/groups and assign roles (View, Edit, Owner).
   b. Storage
      i. Assign one or more vSphere storage policies.
   c. Resource Limits
      i. Set CPU, memory, and storage limits.
   d. VM Classes
      i. Select VM classes to make available (e.g. Small, Medium, Large)
   e. Associate the namespace with a TKR (Tanzu Kubernetes Release) content library.
3. Review settings and create the namespace.
4. Deploy Cluster.

There are multiple ways to deploy a vSphere Kubernetes Service Cluster (Tanzu). If you are running VCF on VMware Private AI Foundation with Nvidia, you can simply create a namespace and cluster using VMware Cloud Foundation Automation. This is an option if you are deploying GPU workloads as the automation will take care of GPU configuration for you. You can also create the Cluster via vSphere client, if the Local Consumption Interface Supervisor Service is installed as well as CLI via "kubectl".

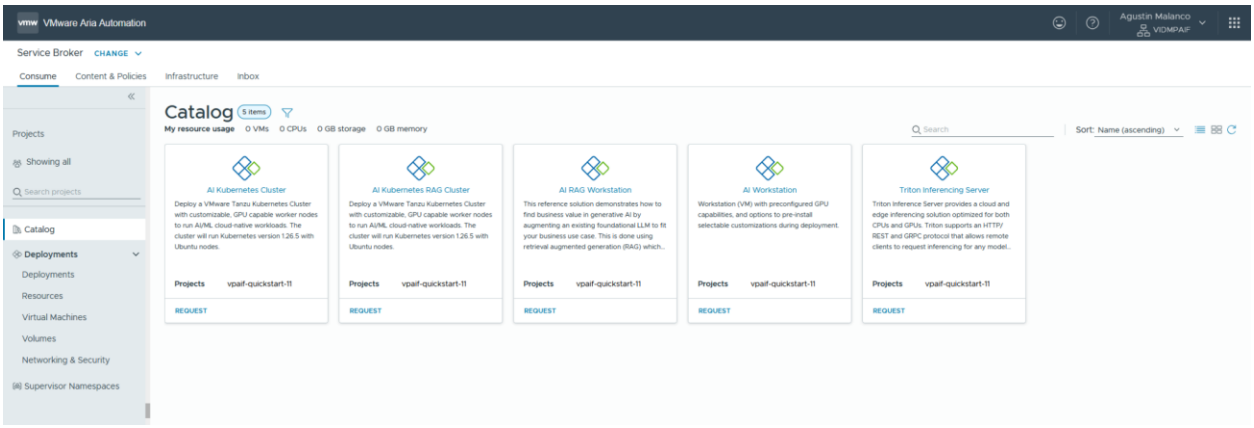


Figure 2: Deploying Cluster via VMware Cloud Foundation Automation Catalog

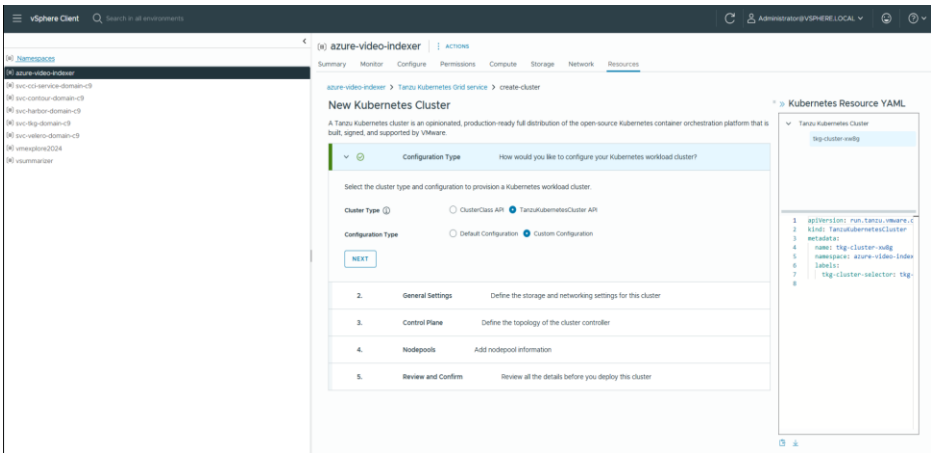


Figure 3: Deploying Cluster via vSphere, Workload Management Section

Example - Deploying TKC via CLI

1. Create your cluster definition file (YAML). See example below:

```
#deploy-azure-vi-cluster.yaml
apiVersion: cluster.x-k8s.io/v1beta1
kind: Cluster
metadata:
  name: azure-vi-cluster
  namespace: azure-vi
spec:
  clusterNetwork:
    services:
      cidrBlocks: ["198.51.100.0/12"]
    pods:
      cidrBlocks: ["192.0.2.0/16"]
    serviceDomain: "cluster.local"
  topology:
    class: tanzukubernetescluster
    version: v1.28.8---vmware.1-fips.1-tkg.2
    controlPlane:
      replicas: 1
      metadata:
        annotations:
          run.tanzu.vmware.com/resolve-os-image: os-name=ubuntu
    workers:
      machineDeployments:
        - class: node-pool
          name: cpu-pool
          replicas: 2
          metadata:
            annotations:
              run.tanzu.vmware.com/resolve-os-image: os-name=ubuntu
          variables:
            overrides:
            - name: vmClass
              value: azure-vi-large
    variables:
      - name: vmClass
        value: guaranteed-large
      - name: storageClass
        value: paif-n
      - name: defaultStorageClass
        value: paif-n
      - name: nodePoolVolumes
        value:
          - name: containerd
            capacity:
              storage: 50Gi
            mountPath: /var/lib/containerd
            storageClass: paif-n
          - name: kubelet
            capacity:
              storage: 25Gi
            mountPath: /var/lib/kubelet
            storageClass: paif-n
```

2. Connect to the namespace created in the previous steps.

   kubectl vsphere login --server=x.x.x.x --tanzu-kubernetes-cluster-namespace azure-vi --vsphere-username user@domain --insecure-skip-tls-verify

3. Deploy the cluster. This process takes about 5-10 min.

   kubectl create -f deploy-azure-vi-cluster.yaml

Prepare Kubernetes Cluster

Before connecting the K8s cluster to Azure Arc and installing the Video Indexer Extension, it is required to:

1. Connect to your newly created cluster. There are different ways to do that:
   a. Using vSphere Plugin

      kubectl vsphere login --server=x.x.x.x --tanzu-kubernetes-cluster-name azure-vi-cluster --tanzu-kubernetes-cluster-namespace azure-vi --vsphere-username user@domain --insecure-skip-tls-verify

   b. Downloading Cluster kubeconfig file from vSphere. This requires the Local Consumption Interface Supervisor service.

      kubectl --kubeconfig kubeconfigfile -parameters

2. Deploy nginx ingress controller then retrieve ingress controller external IP address or DNS name. The controller external IP/DNS address will be used later on for the Video Indexer endpoint URI. You can use other ingress options if available, such as VMware AVI Load Balancer.
3. Provision the following namespaces within TKC with the proper labels.
   a. azure-arc
   b. azure-arc-release
   c. videoindexer

Example - Deploying ingress nginx and provisioning required namespaces

```
#Deploy nginx controller and retrieve External svc IP Address
kubectl apply -f https://raw.githubusercontent.com/kubernetes/ingress-nginx/controller-v1.11.2/deploy/static/provider/cloud/deploy.yaml

kubectl -n ingress-nginx get svc

#Provision required namespaces
kubectl create namespace azure-arc
kubectl create namespace azure-arc-release
kubectl create namespace videoindexer

#Apply necessary labels and annotations to the namespaces
kubectl label namespace azure-arc app.kubernetes.io/managed-by=Helm kubernetes.io/metadata.name=azure-arc pod-security.kubernetes.io/enforce=privileged

kubectl annotate namespaces azure-arc meta.helm.sh/release-name=azure-arc meta.helm.sh/release-namespace=azure-arc-release

kubectl label namespace azure-arc-release kubernetes.io/metadata.name=azure-arc-release name=azure-arc-release pod-security.kubernetes.io/enforce=privileged

kubectl label namespace videoindexer kubernetes.io/metadata.name=videoindexer pod-security.kubernetes.io/enforce=privileged
```

Connecting On-Prem Kubernetes Cluster in Azure Arc

Follow Microsoft's Quickstart guide to connect your on-premises Kubernetes cluster or TKC to Azure Arc.

Create an Microsoft Video Indexer Account

Azure Video Indexer account is a prerequisite for deploying the Video Indexer Extension. The best way to create the account is via Azure Portal. The details can be found here.

Deploying Microsoft Video Indexer Extension

To deploy the Video Indexer Extension, you will need to:

1. Connect to the On-Prem Cluster using AzureCLI.
2. Create Cognitive Services Resources for the extension.
3. Create Azure Arc Video Indexer Extension using CLI. This step requires the ingress-nginx svc external IP address configured previously.

```
az k8s-extension create --name videoindexer \
    --extension-type Microsoft.videoindexer \
    --scope cluster \
    --release-namespace ${namespace} \
    --cluster-name ${connectedClusterName} \
    --resource-group ${connectedClusterRg} \
    --cluster-type connectedClusters \
    --auto-upgrade-minor-version true \
    --config-protected-settings "speech.endpointUri=${speechUri}" \
    --config-protected-settings "speech.secret=${speechSecret}" \
    --config-protected-settings "translate.endpointUri=${translateUri}" \
    --config-protected-settings "translate.secret=${translateSecret}" \
    --config-protected-settings "ocr.endpointUri=${ocrUri}" \
    --config-protected-settings "ocr.secret=${ocrSecret}" \
    --config "videoIndexer.accountId=${viAccountId}" \
    --config "videoIndexer.endpointUri=${dnsName}"
```

Figure 4: Azure Video Indexer Extension deployment parameters

Please consult Microsoft's Video Indexer Documentation for more details regarding the implementation.

Accessing Video Indexer, Uploading Videos and Getting Insights

Finally, we can connect to https://www.videoindexer.ai, switch to the appropriate account, and begin uploading videos for analysis.

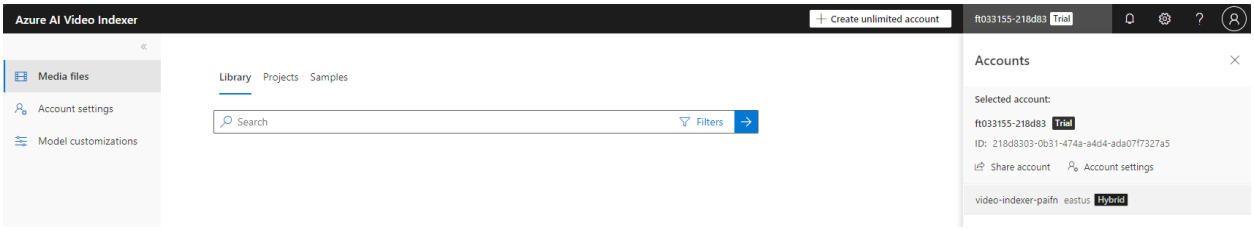


Figure 5: Selecting Video Indexer Account

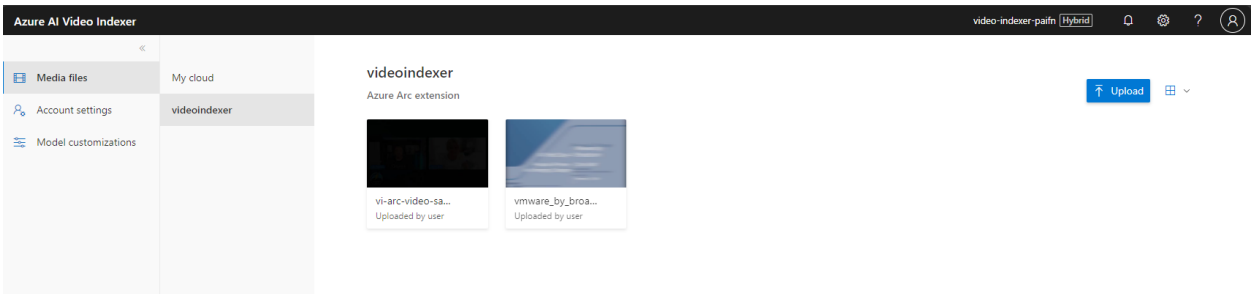


Figure 6: Accessing On-prem Video Indexer Extension

Note: If you are using self-signed certificates, you might need to open link https://ingress-nginx-externalIP/info before connecting to the videoindexer section within https://www.videoindexer.ai. This will avoid videoindexer page not loading correctly (Error: Couldn't reach the Arc extension). Also, you must be able to reach the Azure Video Index endpoint from where you are loading the videoindexer.ai page. See more troubleshooting information here.

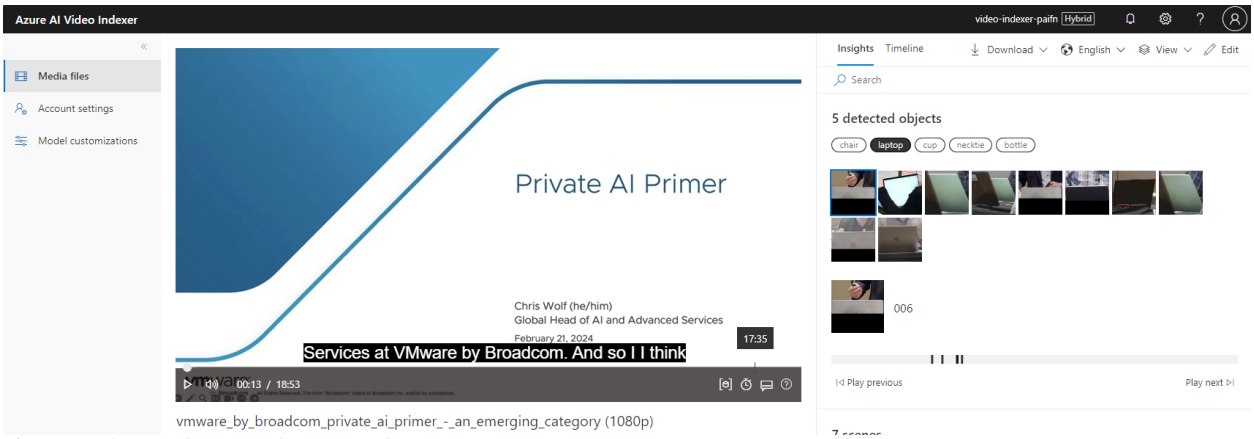


Figure 7: Video Insights - Object Detection

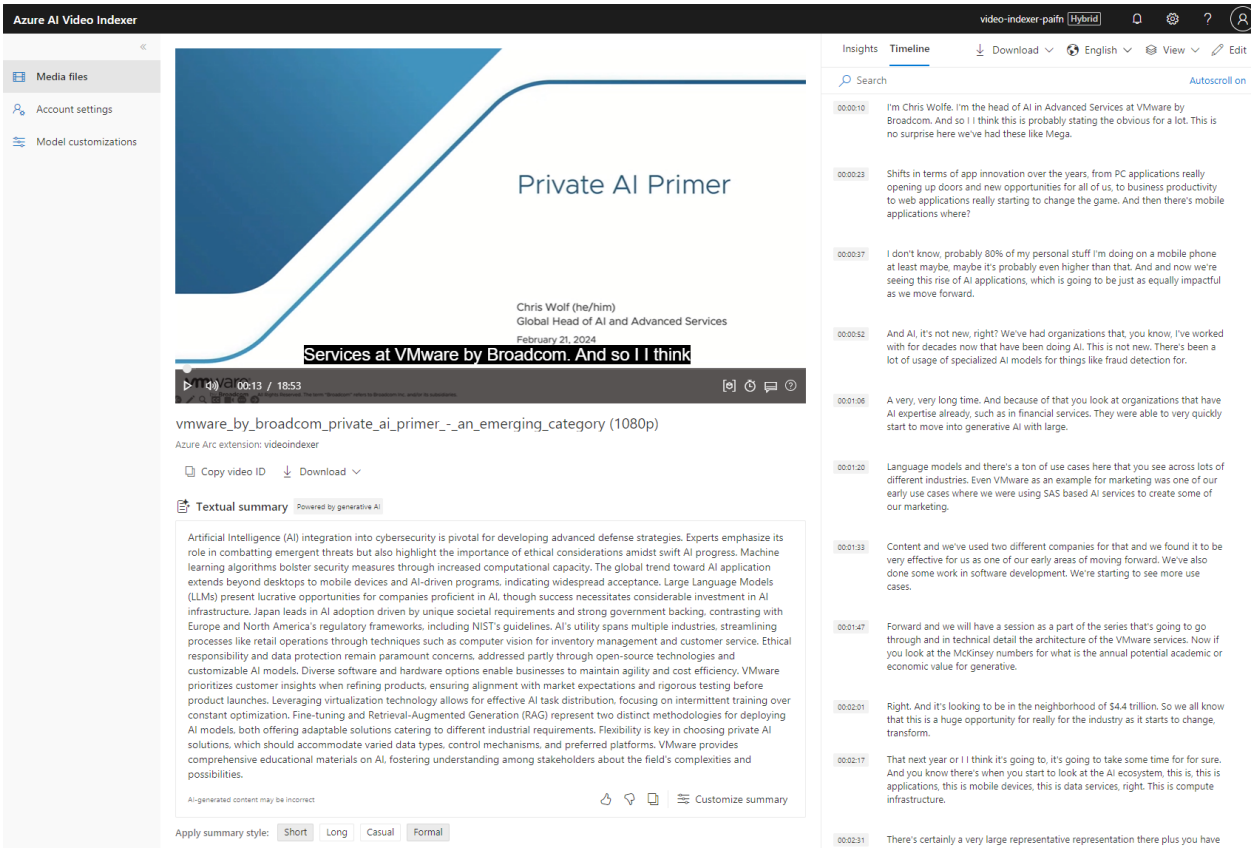


Figure 8: Video Insights - Timeline and AI Generated Textual Summary

Conclusion

The integration of VMware Cloud Foundation and Microsoft Azure Video Indexer, enabled by Azure Arc, represents a significant leap forward in hybrid cloud capabilities and secure AI implementation. This solution addresses the critical needs of modern enterprises by providing a robust platform for running advanced Azure workloads on-premises within the secure confines of VMware's Private AI framework.

By leveraging this integrated solution, organizations can confidently harness the power of AI-driven video analytics while maintaining strict control over their data. This approach not only enhances security and compliance but also drives innovation, operational efficiency, and cost-effectiveness. As businesses continue to navigate the complexities of digital transformation, this VMware-Microsoft collaboration offers a clear path forward, enabling them to stay competitive in an increasingly data-driven world.

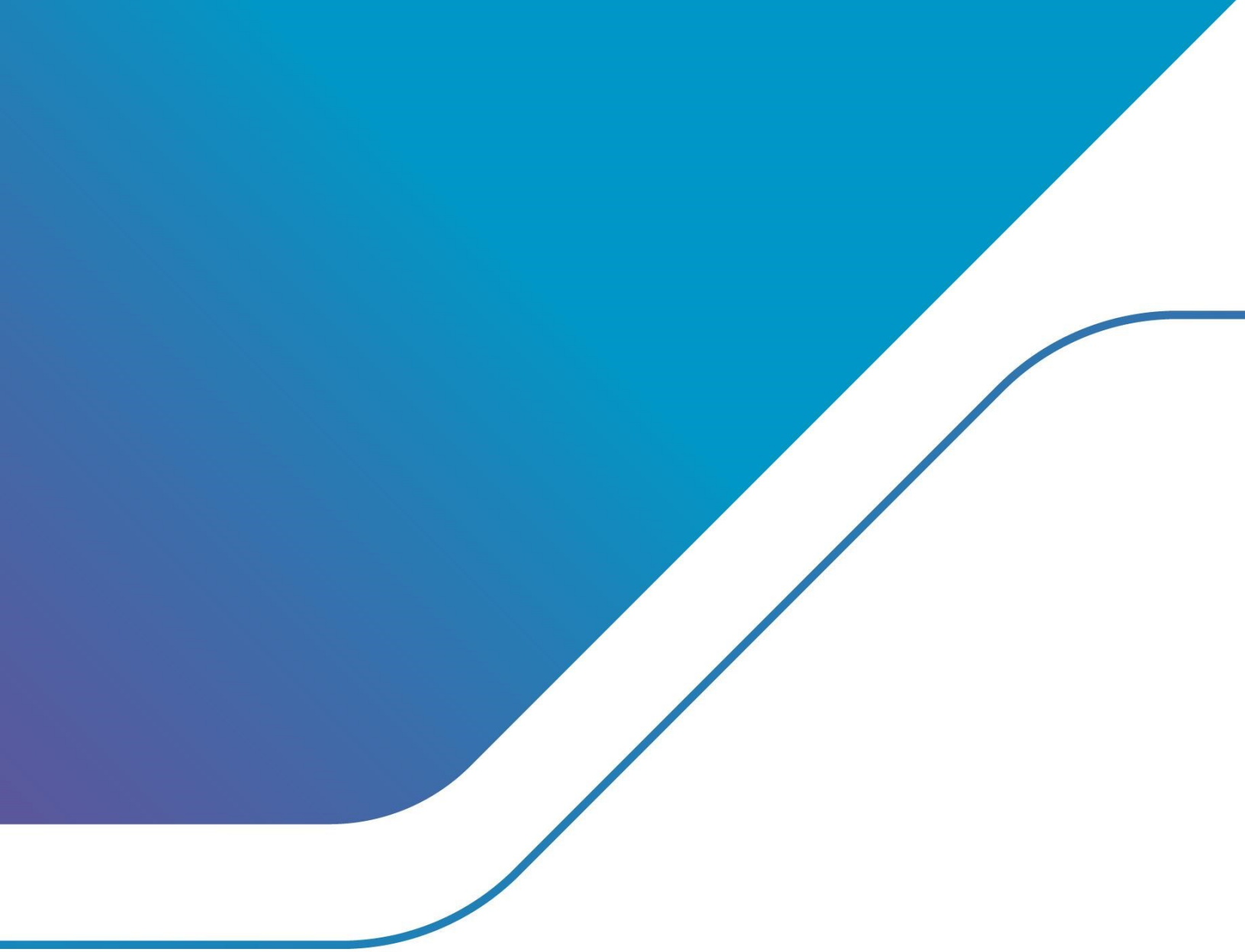
Ultimately, this integration empowers organizations to unlock the full potential of their video content, drive deeper insights, and make more informed decisions—all while operating within a secure, flexible, and familiar IT environment. As we look to the future, it's clear that solutions like this will play a pivotal role in shaping the next generation of enterprise IT strategies, paving the way for a more agile, intelligent, and secure digital landscape.

References

- VMware Cloud Foundation | Private Cloud Platform
- VCF AI/ML Solutions | VMware
- Azure Arc – Hybrid and Multi-Cloud Management and Solution
- Microsoft Azure AI Video Indexer - Unlock Video Insights
- What is Azure AI Video Indexer enabled by Arc? | Microsoft Learn
- Video Indexer Arc Extension

About the Author

- Fabiano Teixeira, Solutions Architect, VMware by Broadcom



vmware by Broadcom

Item No: vmw-bc-wp-tech-temp-uslet-word-2024 1/24