Deloitte.

HYBRID CLOUD

Designing Hybrid Cloud Architecture for the Future

By Doug Bourgeois and Sean VanDruff

Cloud environments are becoming broadly implemented in the government space as organizations look to both gain costs savings and enhance capabilities through digitizing their infrastructure. When properly designed and managed, cloud infrastructures provide greater access to future innovation through enabling next generation services like Artificial Intelligence (AI) and the Internet of things (IoT). As government and industry continue to explore how to best implement the cloud, it is becoming clearer that hybrid cloud (consisting of a combination of on-premise private clouds and external public cloud services) is likely necessary to satisfy the diverse requirements and demands of a next generation system architecture by leveraging the benefits of each type of cloud deployment. As a result, hybrid cloud is anticipated to grow at a CAGR of 20%, resulting in a $4.5B market for the U.S. Government by 2025[1]. It is worth noting that this estimate also includes multi-cloud architectures (the combination of two or more public, external cloud providers) which can optimize application deployments on an even more granular scale and represent the next evolution in cloud environments. (See Exhibit 1.)

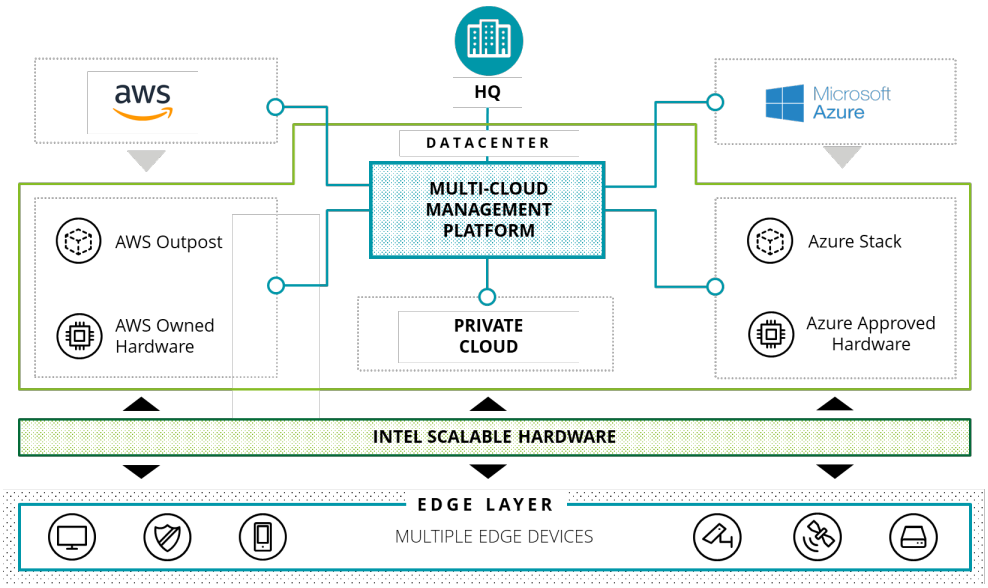
The hybrid cloud architecture of the future must be driven by the missions, business outcomes, and characteristics of next generation workloads. As a result, underlying infrastructure and management frameworks cannot be treated as a commodity, but rather require intentional decisions about the types of hardware systems underpinning cloud environments in order to support the evolving and expanding need for intelligent, cost efficient workloads. This analysis includes selecting the deployment location (public-, private-, hybrid-, or multi-cloud) and specific hardware that enables the mission capabilities.

Due to the level of strategic complexity and advanced technological expertise required to translate mission

[1] Allied Market Research – Global Hybrid Cloud Market Opportunity Analysis and Industry Forecast, 2018-2025

capabilities to an appropriate architectural infrastructure, organizations across industries have recognized that successfully designing and implementing such large scale transformations requires cooperation between multiple industry partners. This paper includes a variety of perspectives gleaned from our work with organizations at all levels of the hybrid cloud stack. While the major Cloud Service Providers (CSPs) and their offerings are generally well known in the marketplace, hardware layer technologies are frequently overlooked. In response to this trend, and as a result of their unique scalable hardware approach, components produced by Intel® are used to illustrate cost and capability comparisons.

EXHIBIT 1 | MULTI CLOUD MANAGEMENT PLATFORM REFERENCE ARCHITECTURE



Reducing Complexity in Managing Hybrid/Multi-Cloud Environments

Many business drivers are pointing to a future of complex and highly specialized system architectures that make use of many diverse cloud deployments to achieve organizational missions. As this landscape continues to mature, Government IT leaders will require an integrated multi-cloud management framework designed to optimize and integrate these diverse, and separate entities into a seamless cloud capability infrastructure. Ultimately this supports a variety of hybrid-cloud mission objectives, including:

- Streamlined, full lifecycle management of all cloud deployed resources and services (On-Premises, Cloud, Edge, and ultimately hybrid multi-cloud) from a single management plane.
- Cost-effective empowerment of IT teams to provide next-gen services (AI/ML/DL/IoT/etc.).
- Full-stack security, configuration attestation, enhanced encryption, key management and other advanced security functions.

To support a government IT mission that leverages hybrid cloud environments and supports next-generation workloads, Deloitte has developed an integrated multi-cloud management (iMCM) solution – leveraging solutions from leading technology companies including Dell EMC, VMWare, Intel Corporation, and other ecosystem participants– which allows application workloads to be deployed to either external clouds or on-premise environments based on requirements and best-fit, while being managed through a common and standardized control plane. This allows for the application of standardized policy, processes, and governance across multiple

cloud environments to help maintain compliance with evolving government mandates without the need to learn the diverse toolsets specific to each cloud model. Consequently, the multi-cloud solution can enable quicker deployments, accelerate security accreditation, and streamline the adoption of cloud services to deliver maximum value to any IT organization.

iMCM is just one example of a broad marketplace of solutions that can help manage complex and broad cloud ecosystems, however, any selected multi-cloud management platform should be ready to incorporate the ongoing advancements to the hybrid cloud market.

Notable Advancements in the Hybrid Cloud Market

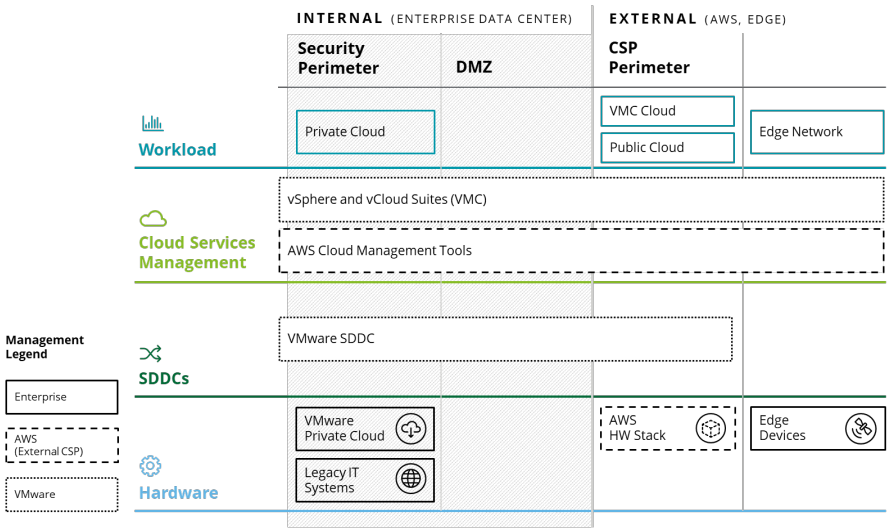
Any multi-cloud management system will need to incorporate diverse cloud providers and make use of the advancements in the hybrid cloud market today. While traditional data center "private clouds" and the Software-Defined Data Center (SDDC) are established concepts in industry, the need to adopt hybrid cloud infrastructure to support next-generation workloads in latency sensitive, bandwidth intensive, and/or data transfer-heavy applications is driving the evolution of the private cloud to more easily integrate with public cloud infrastructure. As a result, public cloud vendors are leveraging the convenience of their capabilities and extending into the on-premises data center and IT infrastructure. Critically, this means that each of these hybrid cloud offerings has similar but different approaches to service delivery in the hybrid cloud model. These differences range from the level and control of management responsibility to where the services reside in respect to the customer's network. The selection of the appropriate offering is a key design decision that will determine the growth and modernization strategy that underpins an enterprise level cloud journey. Notable strategic hybrid offerings from public cloud vendors are included below.

VMware Cloud on AWS (VMWare + AWS Public Cloud)

Since many government private clouds are based on VMWare's software defined datacenter (SDDC), this hybrid cloud offering extends the VMware SDDC to support AWS public cloud infrastructure. By leveraging existing, familiar VMware tools such as vSphere, NSX, and VSAN into the cloud, government IT leaders maintain consistency of operations, and simplify many of the technical challenges involved with managing multiple clouds.

VMC on AWS (See Exhibit 2.) is targeted at organizations with significant footprint and strong operational understanding of VMware's virtualization and software defined data center (SDDC) technologies. This allows these customers to expand into public cloud services using VMWare-centric tools familiar to their technical staff, while also moving toward taking advantage of other AWS native IaaS and PaaS services. It may also be attractive to customers who have recently made significant capital investments in on-premises infrastructure but wish to leverage the capabilities of the AWS cloud platform inside of their security perimeter, while maintaining their VMWare-based management framework. This allows such customers to "set the stage" for an eventual AWS migration while continuing to capitalize on their VMWare-based on-premises environments and existing skillsets. From a service delivery, operations, and management perspective, VMC on AWS is delivered and maintained to the operational SDDC level. The customer need only manage the deployment and usage of the workloads. The VMC on AWS environments tie into their management tools in the VMware vSphere and vCloud Suites allowing the customer's technical teams to manage workloads across the VMC resources. Additionally, it enables customers to migrate between on-premises and external VMware cloud resources.

EXHIBIT 2 | VMWARE CLOUD ON AWS REFERENCE ARCHITECTURE

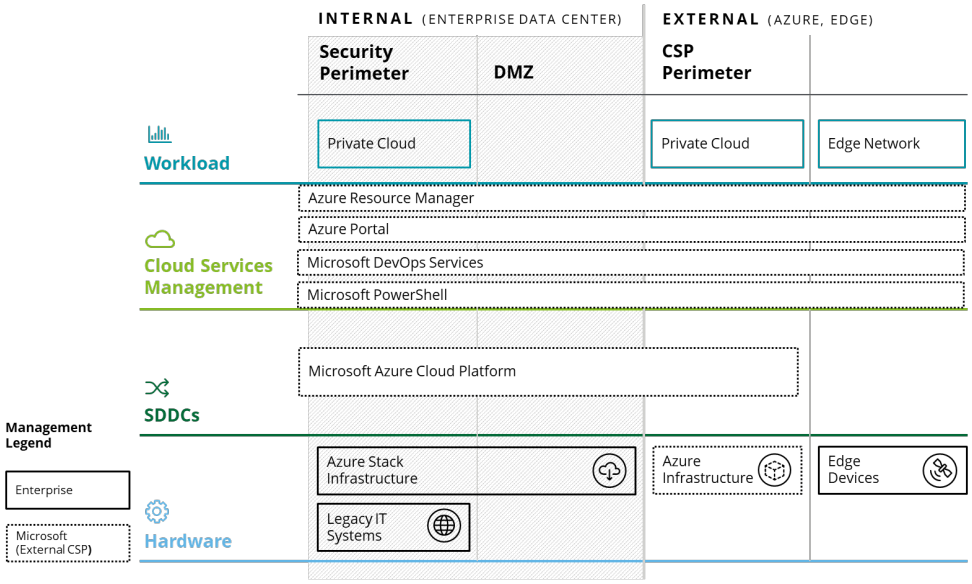


Microsoft Hybrid Cloud (Azure Stack + Microsoft Azure Public Cloud)

Microsoft's hybrid cloud strategy (See Exhibit 3.) seeks to provide a consistent hybrid cloud platform for developers and IT administrators by facilitating the use of familiar Azure tools such as Portal, PowerShell, DevOps, and Azure Resource Manager across cloud boundaries and on the edge.

With a Microsoft-focused administrative team and processes, Azure Stack may be a good fit for customers who have significant investment in Microsoft's virtual ecosystem, such as Microsoft System Center and Hyper-V. Additionally, it may be the right choice for customers who have strong relationships with hardware vendors supporting Azure Stack, validated hardware, and/or are comfortable making continued capital investments in on-premises infrastructure. In such an environment, customers may extend the Azure Platform-as-a-Service (PaaS) into their on-premises environment within their security perimeter and manage the environment using familiar Microsoft tools such as the Administration Portal or PowerShell, while gradually following a migration path to cloud/hybrid cloud infrastructure and the associated cloud-native toolsets.

EXHIBIT 3 | MICROSOFT HYBRID CLOUD REFERENCE ARCHITECTURE



Deloitte.

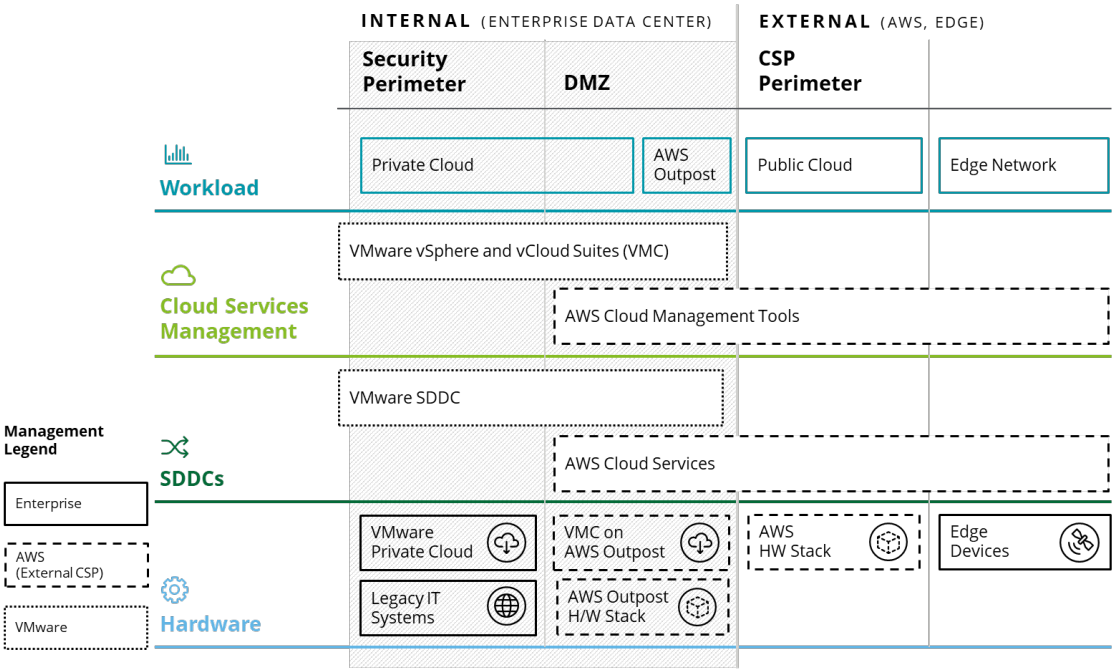
AWS Outposts (AWS Public Cloud Infrastructure On-Premises)

For environments already managed with AWS-trained staff, this hybrid cloud deployment leverages fully-managed and configurable compute and storage racks built with AWS-designed hardware. This allows for on-premises computer and storage workloads while seamlessly using the same native AWS APIs used in the AWS Cloud.

This solution allows for two options: A stand-alone AWS Outpost or an AWS Outpost leveraging VMware Cloud on AWS. A stand-alone Outpost is an extension of AWS services to the client's data center but residing outside of the customer's firewall, acting much like a Demilitarized Zone (DMZ) to the customer's internal data center network. Outposts are fully managed by AWS engineers and consumed by the customer in the same manner as services in AWS data centers (i.e. resources are viewable through the standard AWS portal).

The VMC on AWS Outpost (See Exhibit 4.) option is similar but can reside within the customers network perimeter with VMC resources connected to the customer's existing VMware tools for workload deployment, management, and visibility. Similar to VMC on AWS, this architecture may be appropriate for customers with a significant VMware footprint and strong operational understanding of VMware's virtualization technologies but who do not wish to manage the SDDC stack. Similarly, customers who are approaching an infrastructure refresh and wish to avoid significant capital expenditures (CapX) in favor of moving to an operational (pay-as-you-go/OpX) expenditure model may benefit from adoption of AWS Outposts. As with the previous hybrid models, customers may leverage existing and familiar tools in the Outpost platform while gradually following a path to hybrid cloud infrastructure and cloud-native toolsets.

EXHIBIT 4 | AWS OUTPOSTS REFERENCE ARCHITECTURE



Google Cloud Platform's (GCP) Anthos

As we look towards the future of what Hybrid Cloud might look like, many developers and cloud architects embrace the application of containers. Containers are a form of cloud-native infrastructure targeted at running lightweight applications and coordinating them to conduct more complex tasks through a process called "orchestration". A key feature of containers is their open source nature, allowing for a "bring your own hardware" approach to hybrid cloud. As this container community grows, focus has been drawn to making portable and scalable containers for on-prem and edge applications.

Google Cloud Platform's (GCP) Anthos offering is an example of a solution utilizing an on-prem containerized architecture for hybrid cloud and multi-cloud solutions. Anthos creates a shared containerized architecture using the open source container orchestration tool Kubernetes. A key feature of containers is their open source nature, allowing for a "bring your own hardware" approach to hybrid cloud. By coordinating Kubernetes on all systems on the network, Anthos allows for portable, elastic, and lightweight infrastructure to be available anywhere. As government IT organizations plan for the future and begin to embrace the concept of container-based application architectures they may want to strongly consider an alternative hybrid strategy such as GCP's Anthos managed offering.

Hardware Considerations in Cloud Migration of Advanced Workloads

Hardware is the core foundation of all types of cloud deployments (i.e., public-, private-, hybrid-, and multi-cloud) and should be intentionally selected and combined based on specific mission capabilities. In traditional private cloud environments this selection was intentional as it represented capital expenses for the organization. That level of intention should continue even when considering public cloud selections in the overall hybrid cloud model. This is the case because some public cloud vendors operate "commodity" hardware, chipsets, and infrastructure technologies to provide their offerings at a reduced cost. While this cost reduction is valued by the end user, it may also add complexity and risk when migrating or developing advanced workloads designed to leverage next-generation hardware in the private cloud. For these next generation workloads, hardware cannot be viewed as a commodity and instead should be analyzed in lockstep with mission requirements. These deliberate decisions about the underlying hardware will enable the development of the required capability and, with proper design considerations, can be optimized to the mission space to maximize performance per watt per dollar.

Leveraging Hardware and Software Tools for Advanced Workloads

In response to the need for Government IT Leaders to support next-generation and advanced workloads such as Accelerated Computer Vision, Battlefield Reconnaissance, and Hardware-aware applications, technology companies such as Intel have developed key technologies and products to speed and optimize workloads across multiple cloud deployment types. (See Exhibit 5.) To succeed in the rapidly evolving world of advanced cognitive and analytic mission requirements, IT organizations planning their hybrid and multi-cloud architecture should consider the types of use cases and associated technologies that may be deployed throughout their cloud environments. Some examples of possible next-generation use cases and the technologies which enable their advanced capabilities include:

EXHIBIT 5 | ADVANCE WORKLOAD USE CASES

Use Case	Technology	Description	Results
Accelerated Computer Vision for Medical Imaging	Intel OpenVINO, Intel Xeon Scalable Processors, Intel FPGAs, Intel VPUs	Intel's Open Visual Inference & Neural Network Optimization (OpenVINO) is an open source product that utilizes Convolutional Deep Neural Networks to enable and enhance deep learning and computer vision workloads. It supports heterogeneous execution across computer vision accelerators (CPUs, GPUs, FPGAs, VPUs,) and maximizes performance for deep-learning workloads.	By combining the OpenVINO toolkit with Intel Xeon scalable processors, deep learning inferences performed on X-rays and CT scans, increased an estimated 188x in throughput on bone-age prediction models.[2]
Model Optimization of Existing AI and Machine-Learning Models	Intel OpenVINO Neural Network Optimization Toolkit, Intel Xeon Scalable Processors, Intel FPGAs,	Leveraging OpenVINO allows for the optimization of existing trained models from various frameworks including TensorFlow, mxnet, and ONNX. The advanced Intel Xeon (Cascade Lake) processors support the BFloat16 (BF16) floating point format. BF16 is faster than current FP16 floating point formats for deep learning and associated workloads. Use of Field Programable Gate Arrays (FPGAs) allow the dynamic reprograming chips to workload requirements. FPGAs enable gains in efficiencies that were	Using Intel distribution of the OpenVINO toolkit to optimize existing public models for video processing has shown almost 8x increase in frames per second (FPS) on visual applications, while use of OpenVINO with Intel FPGAs has shown an almost 20x increase in FPS.

[2] Intel and Philips Accelerate Deep Learning Inference on CPUs in Key Medical Imaging Uses. (2018, August 14). Retrieved from https://newsroom.intel.com/

		previously impossible with previous generation fixed-function GPU chips, and specifically boost low-latency parallel processing of AI and ML workloads.	
Edge Processing of Sensor Data in Short Latency Environments (OR: Accelerated Battlefield Reconnaissance through Edge Processing and Accelerated Computer Vision)	Intel Xeon Scalable Processors, Intel Optane Memory, Intel Optane Drives,	Using 3D Xpoint memory media such as Intel Optane memory and drives greatly speed storage controllers, providing NVMe SSD-level speed to inexpensive mechanical drives. This significantly speeds boot times, data access, and write rates on data sets at a fraction of the cost of storing the data on expensive SSD/NAND storage, while allowing for ready-state boot of systems and devices. Combined with Intel Xeon Scalable processors, these devices could significantly speed edge processing.	By leveraging technology such as Vision Acceleration and Xeon advanced processors, edge devices such as sensors, cameras, and drones can on-board process larger data sets such as high-resolution images or streamed sensor data, and make decisions, without the necessity of returning data sets to a central cloud for processing and decision making
Hardware-Aware Applications, Advanced Security & Full-Stack Encryption	Intel SGX, Intel Xeon Scalable Processors, Trusted Platform Module (TPM)	Intel SGX provides new isolation and tamper detection capabilities to augment traditional runtime protections such as TPM functions with boot-level process isolation and tamper detection. This allows for true 'end-to-end' encryption of workloads by leveraging SGX processor and process-level encryption and isolation to complement other current encryption schemas (ex. data at rest, data in flight).	Leveraging of hardware-assisted trusted execution environments into the application itself enables the protection of data and processes, even if a platform's components (ex. BIOS, OS, or firmware) are compromised. Similarly, additional data confidentiality and rights management enhancements can be enabled by determining access levels based on the existence of specific hardware. In this way, an application may be deployed with varying levels of access to sensitive or mission-critical data depending on the hardware used to run the application.

As these capabilities are created and migrated to the cloud it is necessary for IT leaders to consider the underlying hardware required to achieve the mission. (See Exhibit 6.) For example, an application leveraging Intel SGX, or a machine vision application leveraging Intel Xeon scalable processors with FPGAs may experience significant issues if migrated to public cloud architecture which does not incorporate the same hardware features. Instead, they may need to migrate their workloads to public cloud instances with matching architecture. Similar considerations should be provided for other applications designed for intelligent processing of video, voice and natural language, AI/ML, or similar advanced cognitive workloads leveraging Video Processing Units (VPUs), Programmable Acceleration Cards (PACs), and/or Scalable Processing. For reference, included below are example Azure and AWS instance types leveraging Intel architectures that can be selected when developing new or migrating existing workloads in a distributed hybrid cloud.

EXHIBIT 6 | CSP INSTANCE TYPES LEVERAGING INTEL ARCHITECTURE

	General Purpose	Memory Optimized	HPC Optimized	Storage I/O Optimized	Protect Data In-Use	FPGA Enabled
Features	Scalable Processors, Burst Compute	NVMe SSD, Scalable Processors, 3+ TB RAM	MPI Scalability, NVMe SSD, GPUs, Enhanced Networking	IOPS Optimization, NVMe SSD, High Bandwidth Networking	SGX	FPGA, PCI-E Access, DSP Engine
Chip Family	Haswell, Broadwell, Skylake	Skylake	Broadwell, Skylake	Ivy Bridge, Skylake	Skylake-SP, Cascade Lake	Broadwell, Cascade Lake, Skylake
AWS	M5, T3	X1e, Z1d	P, G, F1	I3, D2, H1	C5	F1
Azure	D-Series	Mv2	HC, H	L	DC	Add-On Service

Deloitte.

Conclusion

The hybrid cloud architecture of the future must be driven by the mission. Integrating these next generation workloads such as AI, ML, and Deep Learning neural networks- with the enterprise requirements for addressing risks, threats, compliance, and governance. To address these challenges, cloud vendors are eager to expand their offering portfolios to include on-premises environments by making their ecosystem and hardware commoditization available in the private cloud. As a result, underlying infrastructure and management frameworks cannot be treated as a commodity, but rather require informed, intentional decisions about the types of hardware systems utilized as well as the types of public cloud instances leveraged to support the evolving and expanding need for intelligent, cost efficient workloads.

Let's Talk

Reach out to our team to request a demo and learn more about hybrid cloud architectures and how iMCM hybrid cloud management platform can help you transform your organization.

Contacts:

Doug Bourgeois
Managing Director
Deloitte Consulting LLP
dbourgeois@deloitte.com
+1.571.814.7157

Sean VanDruff
Senior Technology Fellow
Deloitte Consulting LLP
svandruff@deloitte.com
+1.215.446.4314

Thomas Henry
Senior Manager
Deloitte Consulting LLP
thhenry@deloitte.com
+1.571.388.6529