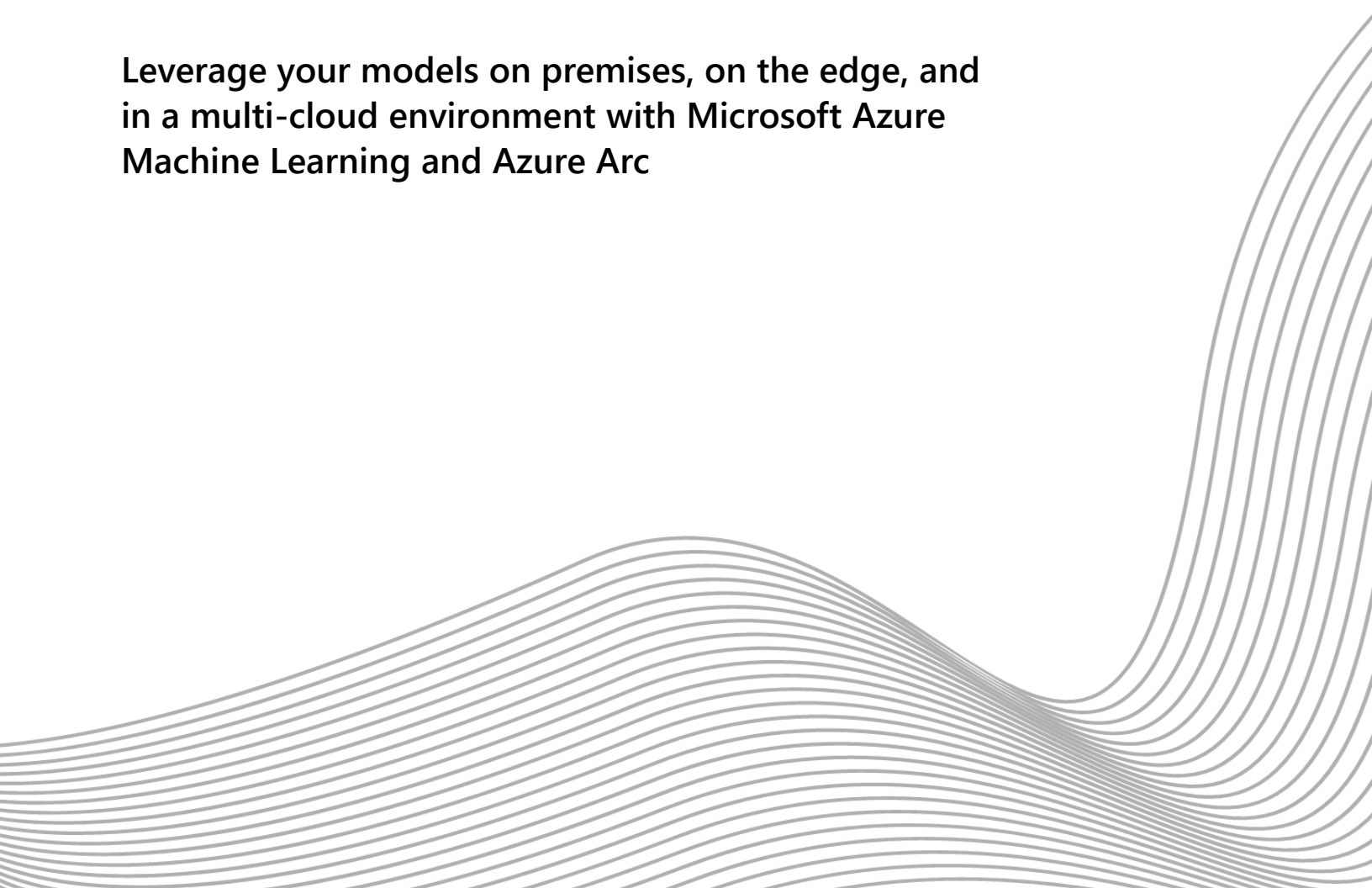# Train and deploy machine learning **anywhere**

Leverage your models on premises, on the edge, and in a multi-cloud environment with Microsoft Azure Machine Learning and Azure Arc

Introduction

Machine learning (ML) has many applications for organizations of all sizes—and the possibilities are constantly expanding. It can help uncover insights in business data, enhance the user experience for customers, and reduce business risk. And while ML can pose unique challenges for organizations, technologies from Microsoft can help your overcome those challenges and harness the benefits of ML.

Azure Machine Learning can help your data team rapidly build and train models and then operationalize them at scale with MLOps. It also lets your data team innovate on an open and flexible platform, all while building responsible ML solutions that protect data and models.

Azure Arc is a set of technologies that extends Azure management and services to any infrastructure. It enables you to build and deploy cloud-native apps and ML models anywhere, at scale. It also provides central visibility, operations, and compliance, and it enables organizations like yours to run Azure services such as Azure Machine Learning anywhere.

Azure Machine Learning integrates with Azure Arc to give you Azure Arc–enabled ML. With Azure Arc–enabled ML capabilities, organizations like yours can make use of Azure Machine Learning anywhere—across one or more clouds, on premises, or the edge.

To show you how these Azure Arc–enabled ML capabilities work, this paper provides examples across use cases and industry verticals, in addition to a case study:

Regulatory compliance: Consumer banking	PII and data sovereignty: Drug discovery	Multi-cloud: Medical imaging	ML on the edge: Renewable-energy infrastructure
Case study: Wolters Kluwer Health			

Contents

The promise and the challenges of ML

ML can bring unprecedented benefits to organizations that embrace it. It can help identify patterns or structures within both your structured and unstructured data, helping you to identify new opportunities. It can deepen your data mining capabilities to find new relationships—and to improve its abilities over time. It can also help reduce business risk by identifying new patterns and tactics to head off fraud.

The challenges to realizing all of the benefits of ML are two-fold: the operations that run ML and the data that powers it. ML operations can pose singular challenges to enterprise IT organizations because traditional IT-management tools are not well-suited for managing ML. At the same time, the data that fuels ML brings unique challenges in terms of quantity and potential sensitivity that can require entirely new processes in order to meet them. Azure Arc–enabled ML is an integration of services and technologies from Microsoft that can supply your organization with the specialized capabilities necessary to realize the benefits of ML.

ML operations require unique tools

ML models have unique characteristics that require specific tools, technologies, and practices to operationalize at scale. Not only must organizations manage the creation of software—the ML models—but the data used to create those models must also be engineered as well. These pipelines must then link up with workflows already in place for solution development.

Once trained and tested, ML models must also be put into production in such a way that they have access to all of their dependencies. Once deployed, the performance of ML models is expected to drift over time as the data used on those models changes. Organizations must continuously monitor their ML models while in production and retrain, repackage, and redeploy them as necessary. Moreover, oftentimes these operations will need to be carried out close to the data for reasons of compliance, latency, or data gravity. For all of these reasons, traditional software deployment practices such as DevOps are insufficient to operationalize ML at scale.

Operationalizing a single ML model can be challenging; deploying multiple models in production at enterprise-scale can be daunting. ML models have multiple software dependencies, and they often need specialized infrastructure for optimal performance. And as models drift or other business metrics for model performance change, ML models must periodically be retrained on new data.

MLOps and Kubernetes

The ML analog to DevOps—MLOps—can streamline these processes across the end-to-end ML lifecycle. MLOps tools, technologies, and processes help operationalize the ML lifecycle by automating many of the steps involved in deploying and managing models in production environments. Enterprises have widely adopted Kubernetes as a container-orchestration platform to host their modern application workloads. Kubernetes can also be used to support MLOps, but in order to enable full ML lifecycle-management on Kubernetes, enterprises have traditionally needed to either write custom code or assemble and install different open-source software packages, which they would then need to manage over the software lifecycle.

ML data challenges

More data almost always yields better ML results. Yet data quantity has a quality all its own, exacerbating challenges that organizations already face for complying with privacy, data-sovereignty, and regulatory requirements, to say nothing of data gravity. To overcome these data challenges, enterprises can incorporate hybrid-cloud and multi-cloud capabilities into their ML practices. However, security, compliance, and intellectual-

property (IP) protection can require isolating ML beyond what public-cloud solutions can provide. Data-gravity and latency considerations can also make on-premises ML attractive for some use cases.

When data teams do make use of cloud resources, a single public cloud–based service rarely meets organizations' needs as the cloud component of a hybrid solution. Organizations might seek to avoid vendor lock-in by being wholly dependent on a single cloud-service provider (CSP), or they might be in an industry vertical like healthcare that relies on integration between multiple organizations on different clouds. For all of these reasons—both operational and data-related—many businesses have thus far foregone implementing ML altogether due to complexity and the need for a simpler way to harness ML for their operations.

Azure Arc–enabled ML

Azure Arc–enabled ML combines the capabilities of Azure Arc and Azure Machine Learning. Azure Arc brings Azure management and services to any infrastructure, providing you with a single pane of glass to manage your entire Kubernetes infrastructure on premises and across clouds. Azure Machine Learning is an enterprise-grade service that enables your data scientists and developers to innovate rapidly using familiar tools and your organization to operationalize ML at scale with MLOps. Azure Machine Learning integrates with Azure Arc to give you Azure Arc–enabled ML. Azure Arc–enabled ML capabilities enable your organization to run enterprise-grade ML workloads on Kubernetes anywhere with the same easy-to-use Azure Machine Learning experience found in the cloud with centralized visibility, operations, and compliance.

Azure Arc

Azure Arc enables organizations like yours to build and deploy cloud-native apps anywhere and at scale. It permits IT to gain central visibility, operations, and compliance on Kubernetes operations and to run Azure services like Azure Machine Learning anywhere. For details about these benefits, see Appendix C. And for more information about Azure Arc, visit its product-overview page: https://azure.microsoft.com/en-us/services/azure-arc/.

Capabilities

Azure Arc–enabled ML enables Azure Machine Learning to provide management for MLOps, rapid innovation for ML, and efficient ML operations on premises, in the cloud, and at the edge.

Unified management

Azure Arc provides single-pane management through the Azure portal. All ML models, no matter where they're built, can be stored and tracked in a central location in Azure Machine Learning for sharing, reproducibility, and audit compliance.

ML is predicated on reproducibility and reusability. This extends from ML models to the whole ML lifecycle. Regardless of its application, ML follows a consistent lifecycle, which is supported at each stage by Azure Machine Learning:

1. **Build**—Data scientists create their initial models and then clean and prepare datasets to be usable by those models.
2. **Train**—With the data prepared, data scientists use it to train models.

3. **Validate**—Once the models are trained, ML engineers create containers for their models to run in with all their dependencies, and the data team evaluates how model performance compares to their business goals.

4. **Deploy**—The data team deploys the models to scale out inference clusters in the cloud or on premises to be used on new data.

5. **Monitor**—As the models are used in production, the data team monitors them for degraded performance and retrains them as necessary to keep them relevant to business operations.
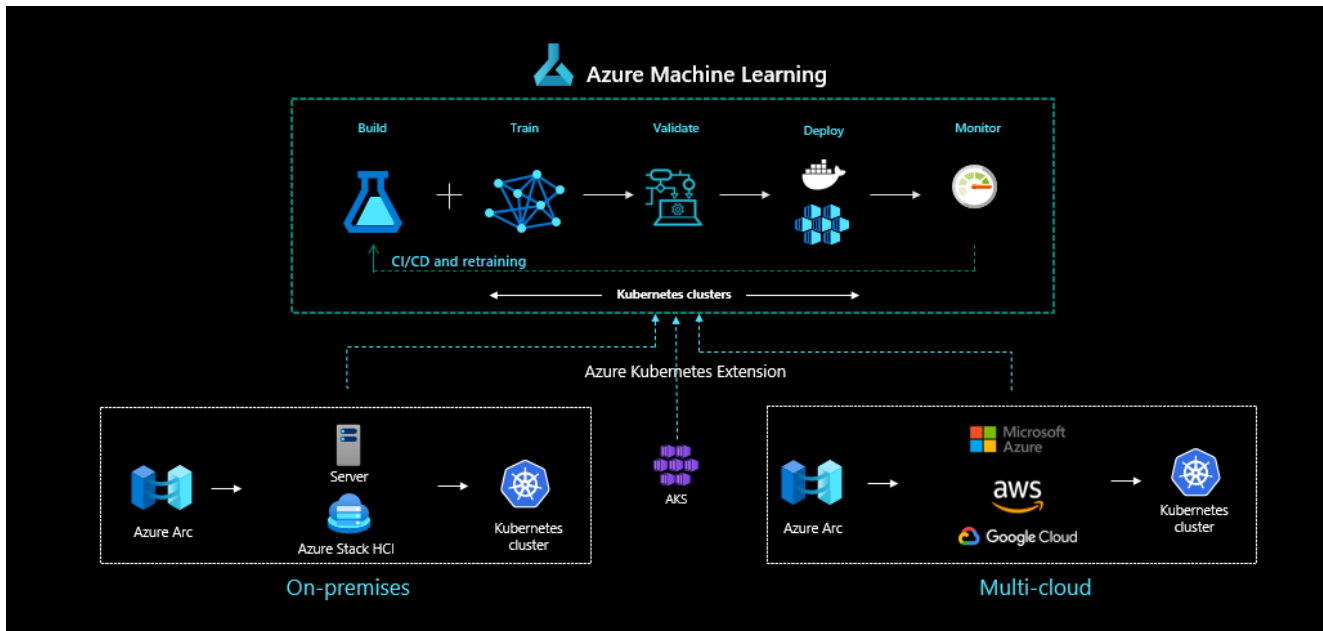

Figure 1. The entire ML lifecycle is managed at all levels by Azure Arc–enabled ML

With whole-of-lifecycle management and tracking, data science can become tenable for organizations of all sizes. With an efficient means of versioning and monitoring the entire ML lifecycle, organizations can head off irreproducible experiments and training runs. This can decrease both the time for the whole ML lifecycle and the chance of errors. Moreover, with automated MLOps, Azure Arc–enabled ML enables organizations to unlock the power of infrastructure based on Kubernetes to run enterprise-grade MLOps.

Rapid innovation

Azure Machine Learning supports open-source tools and frameworks for ML model training and inferencing out of the box. Your data scientists can use familiar frameworks like PyTorch, TensorFlow, or scikit-learn, and they can choose the development tools that best meet their needs, including popular integrated development environments (IDEs), Microsoft Visual Studio Code, Jupyter Notebooks, and command-line interfaces (CLIs), or programming languages like Python and R.

With Azure Machine Learning, your organization can access state-of-the-art responsible ML capabilities to understand, control, and help protect your data, models, and processes. This insight can help you explain model behavior during training and inferencing, and build for fairness by detecting and mitigating model bias. Transparency such as this can help your organization meet regulatory compliance with its ML operations, as can the capability of Azure Machine Learning to preserve data privacy throughout the ML lifecycle with differential

privacy techniques and using confidential computing to help secure ML assets. And Azure Machine Learning automatically maintains audit trails, tracks lineage, and uses model datasheets to enable accountability.

Azure Machine Learning also enables you to auto-scale compute resources for ML. Typical ML-model development starts with development or experimentation on a small amount of data, at which stage a single on-premises instance or a cloud-based virtual machine (VM) is sufficient. However, as you scale up your model training on larger datasets or perform distributed training, Azure Machine Learning creates clusters that auto-scale with each training run to provide sufficient compute resources.

Azure Machine Learning

Azure Machine Learning enables your organization to rapidly build and train models and then operationalize them at scale with MLOps. It also lets your data team innovate on an open and flexible platform, all while building responsible ML solutions that protect data and models. For details about these benefits, see Appendix D. For more information about Azure Machine Learning, visit its product-overview page: https://azure.microsoft.com/en-us/services/machine-learning/.

IT and ML personas and experiences

The IT-operations team and data-science team are both integral parts of the broader ML team. Azure Arc–enabled ML enables the ML team to make use of the advantages of Kubernetes—such as containerization and separation of roles—without the ML infrastructure team or data-science professionals needing to learn Kubernetes. Separating roles makes collaboration easier, as it can be based on different skillsets.

IT operators

IT operators are responsible for setting up the compute resources to be used by the data-science team. To get a Kubernetes cluster up and running for Azure Arc–enabled ML for data-science professionals, IT operators need to complete the following tasks (see Figure 3):

1. Ensure that the Kubernetes cluster is up and running.

2. Connect to Azure with Azure Arc.

3. Deploy the Azure Machine Learning extension to the Kubernetes cluster to support the necessary scenario:

    a. Enable the cluster for model training only.

    b. Enable the cluster for model deployment only (administrators can set up an endpoint with a public/private IP address and Transport Layer Security [TLS] termination proxy).

    c. Enable the cluster for both training and model deployment.

4. Create compute targets in Azure Machine Learning workspaces to which data-science professionals submit workloads, and attach the Kubernetes cluster to the Azure Machine Learning workspace.
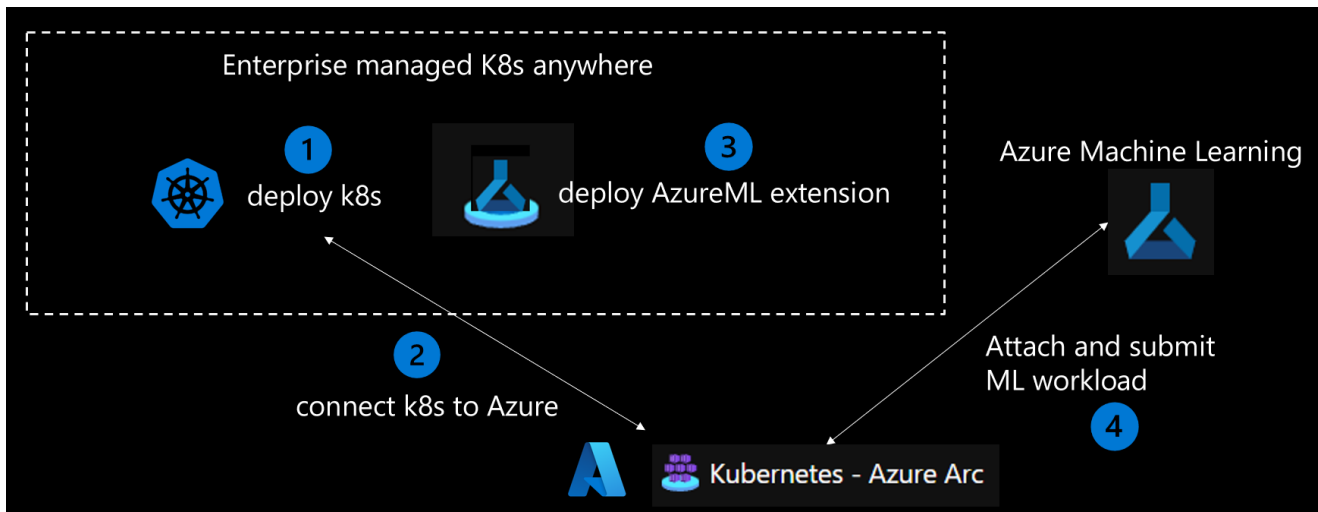
Figure 2. Tasks to deploy and start a Kubernetes cluster for Azure Arc–enabled ML

Administrators can control which engineers and data scientists have access—and they can grant access to entire clusters or selected parts of clusters. ML models and their dependencies can be packaged in discrete Kubernetes containers for deployment without data scientists needing to know Kubernetes.

Data-science professionals

With Azure Arc–enabled ML, data-science professionals can submit ML workloads using compute targets created by IT operators:

- Data scientists can select the tools they like to use. Azure Arc–enabled ML provides different tools, such as using a CLI, the Azure Machine Learning Studio user interface (UI), or the Python SDK (programmatic). The data scientists can then direct those tools to use the compute targets set up by IT operators.

- **Technical note:** To make this process even easier for data scientists and practitioners, they can use Git to clone examples from the Azure Machine Learning repository and then change the compute targets in the examples to their desired compute targets to easily get up and running: github.com/Azure/azureml-examples.

- Data scientists determine which features supported in Azure Arc–enabled ML to use. (For a list comparing the Azure Machine Learning built-in features and features supported in Azure Arc–enabled ML, see Appendix B.)

8

Typical Azure Arc–enabled ML use patterns

Table 1 provides a summary of the typical use patterns for Azure Arc–enabled ML, including where the training data resides in each use pattern, the motivation driving each use pattern, and how the use pattern is realized using Azure Arc–enabled ML.

Table 1. Summary of the typical use patterns for Azure Arc–enabled ML

Use pattern	Location of data	Motivation	Implementation in Azure Arc–enabled ML
Train models in the cloud, deploy models on premises	In the cloud	Make use of cloud compute, either because of elastic compute needs or special hardware in the cloud, such as a graphics processing unit (GPU) Model must be deployed on premises because of security, compliance, or latency requirements	Fully automated MLOps, including training and model deployment steps transitioning seamlessly from cloud to on-premises deployment Repeatable, with all assets tracked properly Model to be retrained when necessary, with on-premises model deployments updated automatically when retrained
Train models on premises, deploy models in the cloud	On premises	Data must remain on premises due to data-residency requirements Deploy model in the cloud for global service access or for compute elasticity for scale and throughput	
ML with network isolation (in cloud or on premises)	In the cloud or on premises	Secure sensitive data or proprietary IP, such as ML models	**In the cloud:** Azure Kubernetes Service (AKS) private cluster with a private link to Azure Machine Learning workspaces; models deployed with private IP address **On premises:** Outbound proxy connection server and Azure ExpressRoute/virtual private cloud connection and private links to Azure Machine Learning workspaces; models deployed with private IP addresses

Note that with the network-isolation use pattern, multiple workspaces share the same Kubernetes cluster. Each business or project team has its own workspace but uses the same Kubernetes cluster to save compute costs and minimize IT management and maintenance overhead.

Example use cases

These examples demonstrate how Azure Arc–enabled ML addresses various customer needs. Where they draw from actual deployments, details have been amalgamated and anonymized.

Regulatory compliance: Consumer banking

Challenge

Banks frequently need to run financial risk-modeling predictions for things like consumer and small-business loans, and they rely heavily on ML models to make lending decisions. While banks' risk models might be trained in datacenters or the cloud, they often need to run ML on premises to meet their regulatory requirements. Banks also need to be able to transparently show regulators and other stakeholders how lending decisions were made through the ML models.

Solution

The hybrid-cloud capabilities of Azure Machine Learning help address these needs. For example, bank risk models can be trained in one or multiple clouds or at a bank datacenter and then deployed on premises to a bank branch. Moreover, the bank can use centralized MLOps to manage the lifecycle of its risk models regardless of where they are trained and deployed. Not only can this help a bank meet its regulatory requirements regarding customer data, it can also enable the deployment of ML models even to bank branches with limited internet connectivity. (For a breakdown of some of the details of using Azure Arc–enabled ML this way, see the *Train models in the cloud, deploy models on premises* and *Train and deploy models on premises* use patterns in Table 1.)

PII and data sovereignty: Drug discovery

Challenge

Pharmaceutical companies must both protect sensitive patient data in drug trials and also conform to data-sovereignty issues as to where trial data can be analyzed. Beyond patient data, firms conducting drug-discovery also want to secure proprietary data concerning candidate drugs and delivery procedures. There can be numerous circumstances in which pharmaceutical companies need to isolate their data, models, and networks. At the same time, data scientists with these firms need to be able to use the same tools for data exploration and model-building, regardless of where datasets reside.

Solution

Because Azure Arc–enabled ML enables models—including existing models trained outside of Azure Arc–enabled ML—to be deployed to almost any environment (be it on premises, on the network edge, or in one or multiple clouds), pharmaceutical firms can train, tune, and deploy their models anywhere, even in isolated, locked-down networks. Doing so can help secure both patient personally identifiable information (PII) and corporate IP during drug research, discovery, and trials. What is more, administrators can still provide comprehensive MLOps to these isolated deployments to maintain efficiency along with security. (For a breakdown of some of the details of using Azure Arc–enabled ML this way, see the *ML in network isolation (in cloud or on premises)* and *Train and deploy models on premises* use patterns in Table 1.)

Multi-cloud: Medical imaging

Challenge

Healthcare organizations commonly need multi-cloud support, given their inter-organizational collaboration structures. Patient data can be scattered between private practices, hospitals, and even different hospital systems—and consolidation within the healthcare industry can result in multiple legacy cloud infrastructures within a single organization. At the same time, healthcare organizations need to secure patient PII and share only necessary details for patient care.

Solution

Using Azure Arc–enabled ML, healthcare organizations can easily perform multi-cloud ML. Models can be trained where patient data is, which can be vastly more efficient in cases using large data files, such as with medical imaging. Models can then be deployed to other locations and organizations can infer medical-images ML models using spare compute on premises or at edge locations and burst to cloud capacity as the need arises. Moreover, ML administrators can still track all of these models to automate the process of deploying retrained models in the future through centralized and automated MLOps. (For a breakdown of some of the details of using Azure Arc–enabled ML this way, see the *Train models in the cloud, deploy models on premises* and *Train and deploy models in one or more clouds* use patterns in Table 1.)

ML on the edge: Renewable-energy infrastructure

Challenge

Wind farms are a great renewable energy source, but their power output fluctuates greatly depending on wind speed and other environmental factors. Wind farms generate petabytes worth of data continuously onsite, and that can be unfeasible to move to the cloud. In addition, ML models can help forecast energy capacity, make power commitments, and facilitate predictive maintenance, but wind farms can have limited or intermittent internet connectivity.

Solution

With Azure Arc–enabled ML, models can be trained on premises without data movement. Wind-energy firms don't have to worry about the economics of shipping data. Models trained at a datacenter or in one or multiple clouds can be deployed to the wind farm and used even if there is no internet connectivity. And automation and MLOps with Azure Arc–enabled ML help ensure that as data at the wind farm changes, the models retrain themselves and maintain accuracy. (For a breakdown of some of the details of using Azure Arc–enabled ML this way, see the *Train models in the cloud, deploy models on premises* and *Train and deploy models on premises* use patterns in Table 1.)

Case study: Wolters Kluwer Health

The Health Division of Wolters Kluwer provides trusted clinical technology and evidence-based solutions that engage clinicians, patients, researchers, students, and the next generation of healthcare providers. The Wolters Kluwer Health data-science team always strives to adopt advancements in technology in building ML models rapidly and efficiently to serve multiple business units within Wolters Kluwer Health. Wolters Kluwer Health enacted modern management for its ML pipeline by re-architecting its ML practice using the primary components of Azure Arc–enabled ML: Azure Machine Learning, Azure Kubernetes Service, and Azure Arc.

Challenge: Isolated training workspaces

The Wolters Kluwer Health data-science team previously used standalone Azure VMs to handle training datasets and ML models. Because manually managing and changing VM tiers based on the varied compute and memory demands of different training runs was cumbersome and impractical, the Wolters Kluwer Health development pipeline used standardized high-compute, large-memory tier VMs, which was costly. In addition, because training jobs were often left to run overnight and there was no practical way to track job completion and automatically shut down idle VMs, Wolters Kluwer Health was forced to consistently pay for unused VM capacity.

Beyond VM expenses, Wolters Kluwer Health had no efficient means of versioning and monitoring its whole model training process and its input/output (I/O) artifacts, which resulted in irreproducible experiments and training runs. The lack of versioned datasets, environments, and models increased development time and multiplied the chance of errors in models. Additionally, whenever the data-science team received updated datasets to retrain models, the entire ML lifecycle had to be rerun manually. This significantly slowed model evaluation and tuning. The Wolters Kluwer Health ML practice was not scalable.

Solution: Azure Machine Learning on Kubernetes

Wolters Kluwer Health got a handle on its ML pipeline by shifting to Azure Machine Learning. The data-science team can now run quick experiments and ad-hoc tasks from within an Azure Machine Learning workspace, eliminating the inefficiencies of performing these tasks on standalone Azure VMs. Wolters Kluwer Health can use workflows in Azure Machine Learning to handle model training and tracking and for automating model retraining. And now, rather than having to set up entire environments again in order to retrain ML models, Wolters Kluwer Health can simply plug in the new data to efficiently handle retraining models.

Wolters Kluwer as a whole standardized its infrastructure on Azure Kubernetes Service (AKS), and the data-science team is able to use a dynamically adaptable cluster model for its model training workflows. And because AKS instance-deployment is handled by Wolters Kluwer IT, the data-science team was able to reduce its self-management overhead. Finally, Wolters Kluwer Health increased ML compute efficiency by using Azure Arc to enable inter-subscription sharing of AKS clusters across Azure Machine Learning workspaces.

Wolters Kluwer Health is now able to track ML assets, with full versioning of all dependent I/O artifacts, along with monitoring of and reproducibility for all training runs. AKS helps Wolters Kluwer Health share the centrally managed compute resources for multiple purposes across the organization, including its ML workflows. And Azure Arc helps Wolters Kluwer Health use the same Kubernetes cluster across multiple ML workspaces provisioned in multiple subscriptions. All of these benefits help the Wolters Kluwer Health data team achieve faster, better results for the many teams across Wolters Kluwer Health that they support.

Enterprise-grade ML for organizations of all sizes and in almost all circumstances

ML poses challenges both because of the unique characteristics of ML models and the surrounding data that powers it. For combinations of legal, logistical, and technical reasons, data cannot always be moved where it would be convenient for training or running ML models. And traditional development and administrative tools are ill-suited to managed ML models, particularly as they proliferate and move between clouds and on-premises locations. Even IT's go-to DevOps solution, Kubernetes, has to be used in new and different ways for ML.

By combining the flexibility and efficiency of Azure Arc with the robust MLOps and operations capabilities of Azure Machine Learning, Azure Arc–enabled ML can help organizations overcome the unique challenges and complexities of ML. Azure Arc–enabled ML simplifies using Kubernetes in ML development and production environments. It also automates many aspects of the MLOps pipeline so that administrators do not have to spend their time wrestling with sprawling ML model deployments and all of their dependencies or with rickety conglomerations of open-source tools. Azure Arc–enabled ML also brings the hybrid capabilities of Azure Arc to ML, meaning that models can be trained in one or more clouds or on premises and then deployed either on premises or to the cloud. This flexibility means that Azure Arc–enabled ML can simplify ML for almost every organization in almost every set of circumstances.

To see Azure Arc–enabled ML in action and learn more about what it can do for your organization, contact Microsoft to set up a proof of concept.

Appendix A: Azure Arc–enabled ML architecture
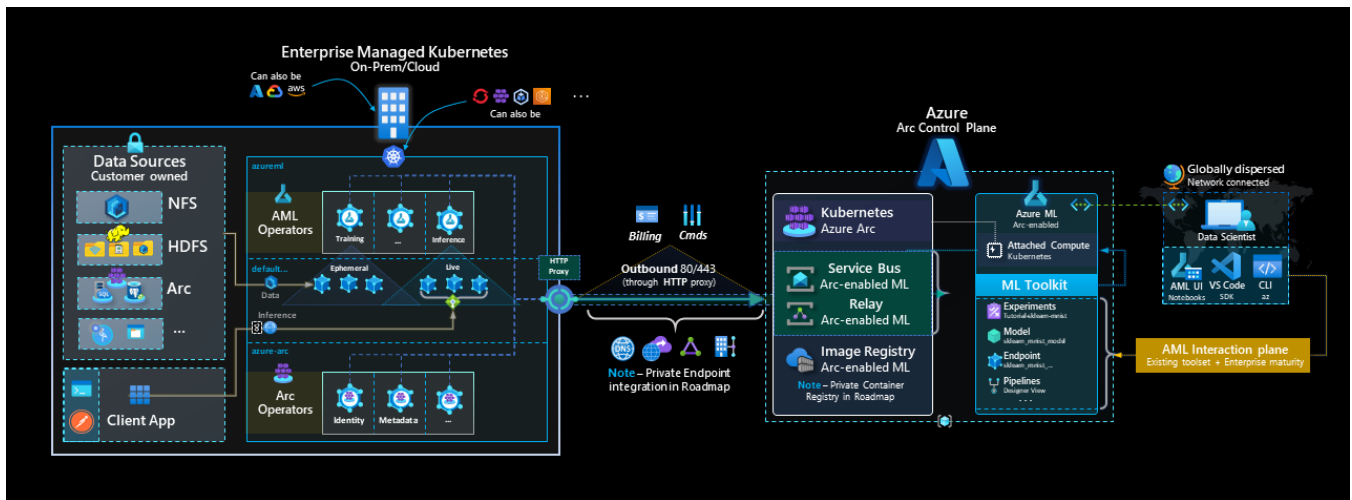
Appendix B: Azure Machine Learning capabilities and Azure Arc–enabled ML

Feature name	Features built into Azure Machine Learning	Features supported in Azure Arc–enabled ML
Train with AzureML Python SDK v1.x	Yes	Yes
Train with AzureML CLI v2	Yes	Yes
Distributed training - MPI, tf.distributed, torch.distributed	Yes	Yes
Train with AzureML pipeline SDK v1.x	Yes	Yes
Train with designer pipeline	Yes	Yes
Deploy with SDK/CLI v1.x	Yes	No
Deploy with CLI v2	Yes	Yes
AKS and Azure Machine Learning workspace cross-subscription support	No	Yes (via Azure Arc)
Multiple workspaces share the same AKS cluster	No	Yes
Node pool targeting for workload	No	Yes
Training with NFS data access	No	Yes
Training with HDFS data access	No	In development
MLflow model integration	Yes	In development
MLOps (GitHub action)	Yes	Yes
Curated image support	Yes	Yes
Safe production rollout with green/blue deployment	Yes	Yes
ML with network isolation	Yes	Yes
Use managed identity to access Azure resources	Yes	Yes
Model-deployment bring-your-own-compute (BYOC) support	Yes	Yes

Appendix C: What is Azure Arc?

Azure Arc is a set of technologies that extends Azure management and services to any infrastructure. Its capabilities bring three main benefits to Azure Arc–enabled ML.

Build and deploy cloud-native apps and models anywhere, at scale

- Seamlessly interact with deployment targets—on premises or public cloud.

- Bring DevOps practices anywhere and build iteratively.

- Reduce errors with consistent policy-driven app and model deployment and cluster operations at scale from source control and templates.

- Save time and reduce admin tasks by using purpose-built Azure data and application services in your architecture.

Gain central visibility, operations, and compliance

- Centrally manage a wide range of resources including Windows and Linux servers, Microsoft SQL Server, Kubernetes clusters, and Azure services.

- Establish central visibility in the Azure portal.

- Meet governance and compliance standards for apps, infrastructure, and data with Azure Policy.

Run Azure services like Azure Machine Learning anywhere

- Get started in minutes with one-click deployment of the managed ML add-on, and train models on any Kubernetes cluster with Azure Machine Learning.

- Take advantage of elastic scale, consistent management, and cloud-style billing models.

- Take ML to environments that might have connectivity constraints, such as edge or Internet of Things (IoT) scenarios.

For more information about Azure Arc, visit its product-overview page: https://azure.microsoft.com/en-us/services/azure-arc/.

Appendix D: What is Azure Machine Learning?

Azure Machine Learning is an enterprise-grade service that enables data scientists and developers to build, deploy, and manage ML models. Its capabilities bring four main benefits to Azure Arc–enabled Machine Learning.

Rapidly build and train models

Accelerate time to value with rapid and accurate model development:

- Improve productivity with Azure Machine Learning studio, the development experience that supports all ML tasks to build, train, and deploy models.
- Collaborate with Jupyter notebooks using built-in support for popular open-source frameworks and libraries.
- Create accurate models quickly with automated ML, using feature engineering and hyperparameter-sweeping capabilities.
- Access the debugger, profiler, and explanations to improve model performance as you train. Use deep Visual Studio Code integration to go from local to cloud training seamlessly, and auto-scale with powerful cloud-based CPU and GPU clusters.

Operationalize at scale

Deploy models with a single click, and manage and govern them efficiently with MLOps:

- Streamline the deployment and management of thousands of models on premises, at the edge, and in multi-cloud environments using MLOps.
- Deploy and score models faster with fully managed endpoints for batch and real-time predictions.
- Use repeatable pipelines to automate workflows for continuous integration and continuous delivery (CI/CD).
- Continuously monitor model-performance metrics, detect data drift, and trigger retraining to improve model performance.
- Enable auditability and governance with out-of-the-box tracking and lineage for all ML artifacts throughout the lifecycle.

Deliver responsible solutions

Understand and protect data and models, build for fairness, and improve model quality:

- Access industry-leading responsible AI capabilities to increase model transparency and improve reliability.
- Understand models using out-of-the-box visualizations and use what-if-analysis to determine feature impact on predictions.
- Share model explanation graphs with your team to ensure compliance. Use state-of-the-art algorithms to test models for fairness issues, compare different models, and take steps to mitigate the issues.
- Identify and debug model errors with the error-analysis toolkit to improve model accuracy.
- Access industry-leading responsible AI capabilities to increase model transparency and improve reliability.

Innovate on a more secure hybrid platform

Run ML workloads anywhere with built-in governance, security, and compliance:

- Increase security across the ML lifecycle with comprehensive capabilities spanning identity, authentication, data, networking, monitoring, governance, and compliance.

- Build more secure ML solutions using custom role-based access control (RBAC), virtual networks, data encryption, private endpoints, and end-to-end private IP addresses.

- Train and deploy models on premises to meet data-sovereignty requirements.

- Manage governance with built-in policies and streamline compliance with a comprehensive portfolio spanning 60 certifications, including FedRAMP High and Health Insurance Portability and Accountability Act (HIPAA).

For more information about Azure Machine Learning, visit its product-overview page: https://azure.microsoft.com/en-us/services/machine-learning/.