

HPE EZMERAL ML OPS

DevOps speed and agility for machine learning

**Build, train and
deploy models
with DevOps-like
speed and agility
at enterprise scale.**

THE CHALLENGES TO OPERATIONALIZING ML MODELS

The ability to apply artificial intelligence (AI) and machine learning (ML) to unlock insights from data is a key competitive advantage for businesses today.

Today’s modern enterprises understand the benefits machine learning can provide and want to expand its use. However, as they attempt to operationalize their ML models, they are running into “last mile” problems related to model deployment and management.

Much like pre-DevOps software development, most data science organizations today lack streamlined processes for their ML workflows, as well as end-to-end analytics processes connecting DataOps and ML Ops, causing many data science projects to fail. Operationalizing ML models is also hard due to the multiplicity of tools and frameworks for data processing, model development, model training, and visualization, as well as numerous security and governance issues.

It may seem like a straightforward solution to use DevOps tools and practices for the ML lifecycle. However, ML workflows are very iterative in nature and off-the-shelf software development tools and methodologies will not work. Public cloud service providers offer disjointed services, and users are required to piece together an end-to-end ML workflow. Also, the public cloud may not be an option for many organizations with workload requirements that require on-premises deployments due to considerations involving vendor lock-in, security, performance, or data gravity.

All of these inhibit model deployment into current business processes and applications.

COMPLETE ML LIFECYCLE COVERAGE

HPE Ezmeral ML Ops is an end-to-end data science solution with the flexibility to run on-premises, in multiple public clouds, or in a hybrid model and respond to dynamic business requirements in a variety of use cases. HPE Ezmeral ML Ops addresses the challenges of operationalizing ML models at enterprise scale by delivering a cloud-like experience, combined with pre-packaged tools to operationalize the machine learning lifecycle.

HPE Ezmeral ML Ops lets customers build, train, and deploy models with DevOps-like speed and agility. It provides a single platform that addresses all aspects of the ML lifecycle—from data prep to model building, training, deployment, monitoring, and collaboration—and operationalizes end-to-end processes across the ML lifecycle, speeding up data model timelines and reducing time to market.

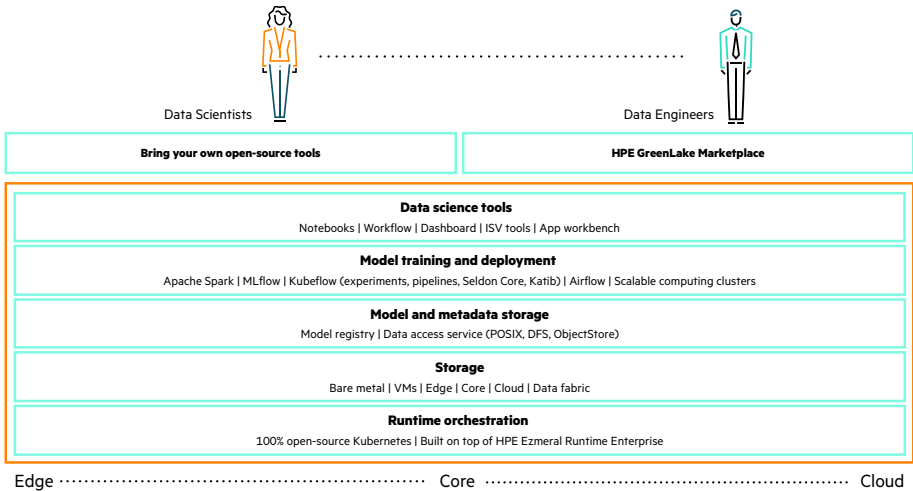


FIGURE 1. HPE Ezmeral ML Ops platform architecture

KEY FEATURES

HPE Ezmeral ML Ops addresses the entire ML lifecycle from data preparation to model building, training, deployment, and monitoring.

TABLE 1. Key features of HPE Ezmeral ML Ops

Model building	Pre-packaged, self-service sandbox environments: Sandbox environments with any preferred data science tools—such as TensorFlow™, Apache Spark, Keras, PyTorch, and more—to enable simultaneous experimentation with multiple ML or deep learning (DL) frameworks.
Model training	Scalable training environments with secure access to Big Data: On-demand access to scalable environments—single node or distributed multinode clusters—to meet workload needs for development, test, or production workloads. Patented innovations provide highly performant training environments—with compute and storage separation—that can securely access shared enterprise data sources on-premises or in cloud-based storage.
Model deployment	Flexible, scalable, endpoint deployment: Flexibility to run workloads where needed—on-premises, in public clouds, or a hybrid model. HPE Ezmeral ML Ops deploys the model's native runtime image, such as Python, R, H2O, into a secure, highly available, load-balanced, and containerized HTTP endpoint. An integrated model registry enables version tracking and seamless updates to models in production. Autoscaling from HPE Ezmeral ML Ops dynamically scales nodes for scoring engines.
Model monitoring	End-to-end visibility across the ML lifecycle: Complete visibility into runtime resource usage such as GPU, CPU, and memory utilization. Ability to track, measure, and report model performance, along with third-party integrations such as ParallelM to monitor model performance in production and track accuracy and interpretability.
Collaboration	Enable CI/CD workflows with code, model, and project repositories: Project repository and GitHub integration of HPE Ezmeral ML Ops provides source control, eases collaboration, and enables lineage tracking for improved auditability. The model registry stores multiple models—including multiple versions with metadata—for various runtime engines in the model registry.
Security and control	Secure multitenancy with integration to enterprise authentication mechanisms: HPE Ezmeral ML Ops software provides multitenancy and data isolation to ensure logical separation between each project, group, or department within the organization. HPE Ezmeral ML Ops integrates with enterprise security and authentication mechanisms such as LDAP, Active Directory, and Kerberos. Access shared enterprise data with complete security and access-control mechanisms in place, eliminating requirement for local copies of data.
Hybrid deployment	On-premises, public cloud, or hybrid: HPE Ezmeral ML Ops runs on-premises on any infrastructure, on multiple public clouds (Amazon® Web Services, Google Cloud Platform™, or Microsoft Azure), or in a hybrid model, providing effective utilization of resources and lower operating costs.

KEY BENEFITS

Faster time-to-value: You can manage and provision development, test, or production environments in minutes as opposed to days; and instantly onboard new data scientists with the preferred tools and languages without creating siloed development environments.

Improved productivity: Data scientists spend their time building models and analyzing results rather than waiting for training jobs to complete. HPE Ezmeral ML Ops helps ensure no loss of accuracy or performance degradation in multitenant environments. It increases collaboration and reproducibility with shared code, project, and model repositories.

Reduced risk: It provides enterprise-grade security and access controls on compute servers and data. Lineage tracking provides model governance and auditability for regulatory compliance. Integrations with third-party software provide interpretability. High availability deployments help ensure critical applications do not fail.

Flexibility and elasticity: You can deploy on-premises, cloud, or in a hybrid model to suit your business requirements. HPE Ezmeral ML Ops autoscales clusters to meet the requirements of dynamic workloads.

LEARN MORE AT
hpe.com/info/mlops

Make the right purchase decision.
 Contact our presales specialists.

