SOLUTION BRIEF

D2iQ Kubernetes Platform and NVIDIA DGX™ Systems

The Winning Combination for AI and Machine Learning on Kubernetes

Overview

Kubernetes is a natural fit for AI and Machine Learning (ML) workloads, as it is well-suited to meet their scalability needs and to embrace their continuous development nature. There are, however, several significant challenges to overcome:

- **High risk:** Up to 87% of machine learning initiatives are abandoned before they reach production.
- **Long time to value:** For those initiatives that do make it to production, it can take more than 3 months for a single model to be deployed. Software provisioning at enterprises can take weeks or even months, which adds time and delays obtaining value.
- **Complexity:** Build-or-buy decisions for scalable platforms require immense knowledge of cloud-native infrastructure as well as the entire ML landscape.

Investments in AI and ML projects that don't make it to production are completely lost and the longer it takes to get models into production, the longer it takes to get to a positive ROI (and the lower that total ROI will be).

Fortunately, the right combination of hardware and software can provide the firm foundation needed to reduce or eliminate these challenges and ensure your AI and ML success.

Combining the purpose-built power and flexibility of NVIDIA DGX systems with the D2iQ Kubernetes Platform (DKP)—certified as part of the NVIDIA DGX-Ready Software program—speeds time to solution, maximizes operational efficiency, and provides the governance, security and compliance capabilities needed for true enterprise grade AI and ML operations.

The Winning Combination

The Hardware

Whether creating quality customer experiences, delivering better patient outcomes, or streamlining the supply chain, enterprises need infrastructure that can deliver AI-powered insights. NVIDIA DGX systems deliver the world's leading solutions for enterprise AI infrastructure at scale.

The Software

The D2iQ Kubernetes Platform (DKP) is a certified solution in the NVIDIA DGX-Ready Software program. DKP consists of D2iQ Kaptain, D2iQ Konvoy, and D2iQ Kommander, plus all the training and consulting services and full-stack support needed to quickly get enterprise customers into production at scale.

Key Benefits

Faster time-to-market with both hardware and software purpose built to get AI/ML workloads into production fast

Built-in and fully tested GPU support to simplify the process of exploiting the power of NVIDIA DGX™ Systems

Elimination of operational barriers between model prototypes and production

Enterprise-grade and ready for Day 2 operations with integrated observability, end-to-end security, and cost management

End-to-end notebook tutorials to speed up onboarding, time to value, and production readiness of ML models at scale

Lowest TCO for end-to-end ML without lock-in, DKP promises compatibility with upstream open source components

End-to-end full stack support for Kaptain, Konvoy, Kommander, and all of the platform applications they provide for Day 2 operations

Flexible deployment options, including highly-secure air-gapped networks

Kubernetes made easy with a simplified, automated installer that provisions and installs a complete, flexible, and opinionated open source Kubernetes distribution with best-in-class supporting services

D2iQ Kaptain is an enterprise-ready end-to-end machine learning platform, powered by Kubeflow, that addresses both structural and organizational ML challenges and accelerates the time to market and positive ROI by breaking down the barriers between ML prototypes and production. D2iQ Kaptain enables organizations to develop and deploy machine learning workloads at scale, while satisfying the organization's security and compliance requirements, thus minimizing operational friction and meeting the needs of all the different teams involved in a successful ML effort.

D2iQ Kaptain is powered by Kubeflow, which is the open-source machine learning toolkit for Kubernetes. Unfortunately, Kubeflow is not a single product: it is a collection of several different projects that are loosely coupled and do not provide a seamless user experience. Most components are tested in isolation, which means it is possible for individual components to break as there is little or no end-to-end testing in the open-source edition. Kubeflow is also not an end-to-end machine learning platform, it is a toolkit: there is no unifying user experience that spans the multitude of components.

D2iQ Kaptain is an opinionated subset of Kubeflow projects (approx. 20+ Kubernetes Operators) with some additional components (like Spark and Horovod) that are not part of the Kubeflow ecosystem but are valuable, enterprise-grade additions. These operators enable full lifecycle support, which is absent from the upstream. All the components curated for Kaptain are subject to D2iQ's rigorous testing procedures, and Kaptain's end-to-end security model provides the security features that enterprises require in production environments and that are lacking in Kubeflow.

D2iQ Kaptain approaches end-to-end machine learning in the enterprise "notebooks-first." This means notebooks are the primary development environment for data science, machine learning engineering, and operations teams. We aim to give data scientists tools they are familiar with to perform tasks that are usually outside of their area of expertise: deployments and operations, as that is where the majority of enterprise efforts fail—in the last mile of ML where a positive ROI is in sight but often out of reach.

D2iQ Kaptain runs on D2iQ Konvoy, a comprehensive, enterprise-grade Kubernetes distribution built on pure open source with the platform applications needed for Day 2 production selected, integrated, and tested at scale, for faster time to benefit, enabling companies to leverage Kubernetes with an easy, enterprise-grade experience out of the box and an accelerated time-to-market for your application development needs.

Key Benefits *continued*

Centralized observability to provide enhanced visibility and control at the enterprise level, with comprehensive logging and monitoring across all clusters

Federated management to ensure roles and responsibilities can be separated to deliver policy-driven control and secure provisioning of services—even on shared, multi-tenant infrastructure

Organizational governance that empowers the organization to govern Kubernetes usage to assist with compliance for regulatory, IP and other unique organizational needs

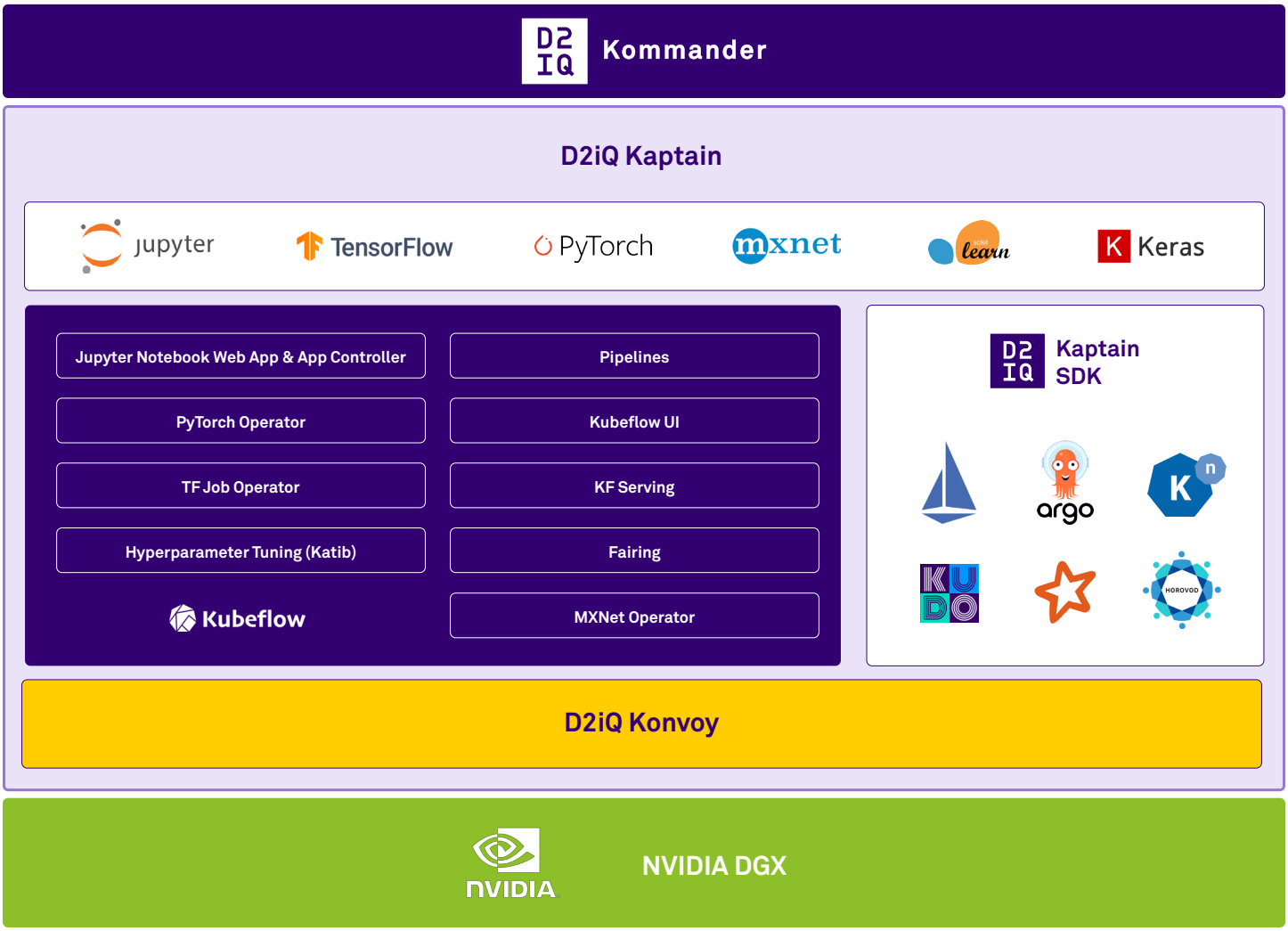
CNCF certified Kubernetes distribution, Service Provider, and Training Partner

Integrated backup and disaster recovery capabilities with Velero

Leverage existing authentication and directory services for secure access and single sign-on to broad cluster-based resources

Zero downtime service upgrades. Simplified and consistent configuration for services and cross cluster operations

And for orchestrating workloads across multiple clusters, D2iQ Kommander is a federated management plane which provides organizations with unified visibility and control over a wide expanse of Kubernetes resources to deliver scale, consistency, governance, and operational efficiency, regardless of distribution, across an organization's on-premise and cloud footprint. D2iQ Kommander simplifies the provisioning of new services and creates greater policy-driven consistency across Kubernetes clusters within the environment. This level of broad control can help organizations meet risk and compliance demands as they govern how and where new application services are used, as well as who is able to engage in policy and operational needs of those services.



Drive Success on Day 2 From Day 0 With D2iQ

The Leading Independent Kubernetes Platform

As you look to leverage Kubernetes as a foundational platform to drive cloud native applications and services, navigating the complexities of the technology will be next to impossible without deep domain expertise and expert support. With this in mind, we designed the D2iQ Kubernetes Platform to provide you with an enterprise-grade Kubernetes foundation, as well as give you access to expert insights, training, services, and support.

Whether you're in need of a guided, opinionated distribution that reduces time to market from months to days, management or operation of your production deployment, or someone to pick up the phone in the middle of the night, D2iQ can deliver a complete spectrum of technology solutions and services to enable an enterprise Kubernetes experience that is ready to meet Day 2 operational expectations.

And as a DGX-Ready Software program partner, D2iQ can help you quickly and easily optimize your use of DGX systems to get high value AI/ML workloads into production at scale. To learn more, contact us.

To learn more about about how D2iQ can be your partner in the cloud native journey, go to www.D2iQ.com.