Accelerate and Operationalize AI/ML and Cloud Native Workloads with WekaIO™, Penguin and Redhat

A Reference Architecture with Weka Limitless Data Platform, Penguin OriginAI and Redhat Openshift Container Platform and Opendatahub

ACCELERATE

Better performance than local NVMe Flash even with containers.

Parallelism to match containers, microservices HPC architecture.

OPTIMIZE

Mixed workload handling.

Resolve poor utilization of expensive GPU resources.

Global namespace with NVMe flash for performance and Redhat Ceph object storage for economics.

SIMPLIFY

Checkpointing for large training jobs.

Sharing training data, logs and inference results.

End to End Security.

Unified management.

ELASTIC AND HYBRID

Built-in data management – protection, DR, cloud-bursting with snap2obj.

Scale as you grow.

Run on-premises or hybrid.

Organizations embark on their Digital Transformation journeys for many reasons, where they are looking to reduce their time to market, improve productivity, reduce risk and launch new business models. As they leverage Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) and cloud native applications, they are increasingly adopting technologies like containers, NVMe flash based storage and GPUs.

For organizations that are new to high performance computing (HPC) and high performance data analytics (HPDA) systems, developing AI infrastructure can be costly and time consuming, taking an estimated two to three years of internal research to bring AI-based products to market. In fact, 78% of AI projects are stalled even before they can be deployed (Dimensional Research, 2019).

Containers became the popular AI/ML deployment model due to their light weightiness, their immutability and portability, and their agility with orchestration provided by Kubernetes, deemed as the cloud native operating system.

KUBERNETES CHALLENGES IN THE ENTERPRISE

80% of enterprises list persistent storage, data management and disaster recovery as top barriers limiting container adoption.

Kubernetes at scale is hard, especially when it comes to data management. Providing stateful-ness is only part of the problem, since the state needs to be protected, needs to provide mobility and disaster recovery across hybrid workflows, when compute and storage are disaggregated. Kubernetes benefits need to be extended over cloud native, enterprise and AI/ML applications alike. Equally important is to provide better TCO and utilization of expensive GPU resources, scale performance and capacity independent of each other, while improving productivity.

With more than 76% of enterprises expected to standardize on Kubernetes within next three years, having a certified, best of breed Origin AI solution with industry leaders – Weka, Penguin computing and Redhat helps customers overcome the challenges of Kubernetes and AI/ML data pipelines and deliver value from the data center to the cloud to the edge (aggregation) , with at-scale performance.

Built by open source leaders, Red Hat® OpenShift® is a leading enterprise Kubernetes platform: a security-focused, consistent foundation to deliver applications anywhere, with streamlined developer workflows to get to market faster. With Red Hat OpenShift, innovators can focus on what matters most: staying competitive and continually exceeding customer expectations.

Red Hat OpenShift has everything needed for hybrid cloud, enterprise container, and Kubernetes development and deployments. It includes an enterprise-grade Linux® operating system, container runtime, networking, monitoring, container registry, authentication, and authorization solutions. These components are tested together for unified operations on a complete Kubernetes platform.

Open Data Hub is a blueprint for building an AI as a service platform on Red Hat's Kubernetes-based OpenShift® Container Platform. It inherits from upstream efforts such as Kafka/Strimzi and Kubeflow, and is the foundation for Red Hat's internal data science and AI platform. Data scientists can create models using Jupyter notebooks, and select from popular tools such as TensorFlow™, scikit-learn, Apache Spark™ and more for developing models. Teams can spend more time solving critical business needs and less on installing and maintaining infrastructure with the Open Data Hub.

WekaIO (Weka) offers WekaFS, the modern parallel file system that is used by eight of the Fortune 50 enterprise organizations to uniquely solve the newest, biggest problems holding back innovation and discovery. Purpose-built to unlock the full capabilities of today's accelerated and agile data center, WekaFS is optimized for NVMe-flash and the hybrid cloud. WekaFS, the world's fastest and most scalable parallel file system for data-intensive workloads, addresses the shareability, performance, and portability challenges by providing stateful and parallel storage, allowing seamless deployment on-premises and easy migration to the cloud. Using the WekaFS Kubernetes CSI plugin, organizations now have increased flexibility in how and where they deploy containers while delivering local storage performance and latency.

PENGUIN COMPUTING ORIGINAL

Penguin Computing partnered with WekaIO, Inc., to create OriginAI TM, a comprehensive, end-to-end solution for datacenter AI that breaks down software, server/switch hardware, data storage and governance, and infrastructure barriers that cause AI projects to fail or stall. A full-service consultancy, Penguin Computing's Analytics Practice acts as a single point of reference for hardware, software, architectural design, hosting, and management, enabling organizations to focus their time and resources on the business and human challenges in bringing AI projects to production. By providing organizations with a comprehensive solution, Penguin Computing radically reduces the time to insight from years to months.

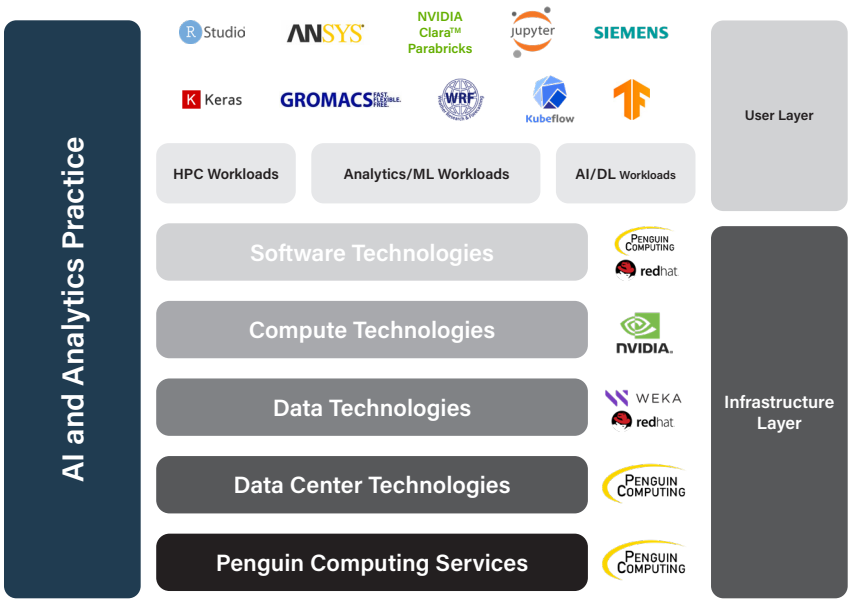
> **The Penguin Computing AI team has tested and validated the Origin AI reference architecture with our partners Weka and Redhat. The Weka Data Platform provides us and other ISV and IHV technology partners with the flexibility and technology innovation to build custom solutions to solve industry challenges through AI.**
>
> ***Kevin Tubbs,*** *SVP, Strategic Solutions Group at Penguin Computing*

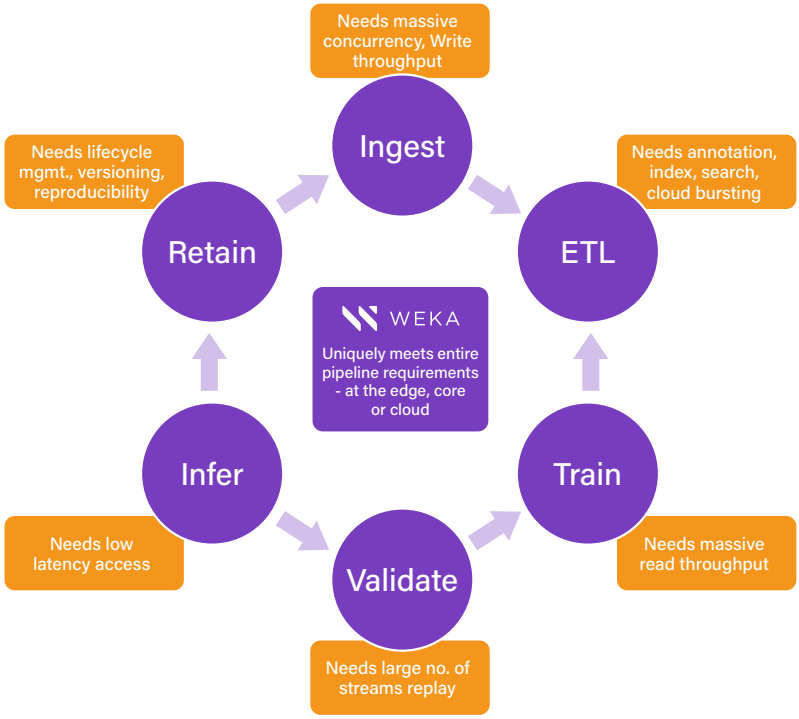
> **By partnering with Redhat and Penguin,, we can offer a solution that fulfills the enterprise- and cloud-ready features required of today's AI/ML container ecosystems. We believe this to be a key step in fulfilling our vision of helping organizations overcome their biggest storage problems with AI- and Cloud-first strategies, so that they can extract more value from their data, faster.**
>
> ***Shailesh Manjrekar,*** *Head of AI and Strategic Alliances, Weka*

WEKA SOLVES STORAGE PERSISTENCE CHALLENGES WITH ITS CSI PLUG-IN

Using the WekaFS Kubernetes CSI plugin organizations now have increased flexibility in how and where they deploy containers, all while delivering local storage performance and low latency. In fact, throughout the system WekaFS delivers the speed-to-market required of an AI-first solution. The WekaFS CSI plugin is deployed using a Helm Chart, along with the POSIX agent on Kubernetes worker nodes. WekaFS supports volume provisioning in both the dynamic (persistent volume claim) and static (persistent volume) forms with its own storage class. It also supports ReadOnlyOnce, ReadOnlyMany, ReadWriteOnce, and ReadWriteMany access modes.

WekaFS provides multi-protocol access to containerized applications with POSIX, NFS, SMB





and S3. WekaFS also provides a unique way of data protection, mobility and disaster recovery in a Kubernetes environment, with its snap2object feature, where a persistence volume claim can match to a virtual filesystem within WekaFS global namespace. WekaFS supports 1024 of these virtual filesystems in a given global namespace, which can extend over NVMe flash tier and private or public S3 bucket.

WekaFS works with Redhat Ceph Storage to provide capacity tier under a single global namespace. Redhat Ceph Storage provides the durability, geo-replication and the economics needed of a capacity tier. WekaFS also provides integration with leading KMS (Key Management Systems) for end to end and at-rest security.

WekaFS, thus, addresses the shareability, acceleration, and portability challenges by providing stateful, parallel storage. WekaFS also allows seamless deployment on premises and easy migration to the cloud, all while meeting the requirements of high performance, mixed-workload handling, and low latency.

To find out more or to arrange for a free trial, contact us at info@weka.io.
To get started right away on your own self provisioned cluster go to https://start.weka.io.

* SPEC® SFS2014 Results: https://bit.ly/35UHLDk
WekaFS™ Data Sheet: https://bit.ly/36BRJdc
Weka AI Data Sheet https://www.weka.io/wp-content/uploads/files/2017/11/Weka_AI_datasheet_W07R4DS201807_Web.pdf

TARGET MARKETS AND USE CASES

ENTERPRISE VERTICAL SOLUTIONS
Banking
Government
Healthcare and Life sciences
Manufacturing and Energy
Telco
Retail

ENTERPRISE USE CASES

HPC and AI/ML MLOps data pipelines
- Autonomous vehicles, NLP/NLU, Lifesciences, Simulations

CI/CD DevOps pipelines
- Cloud native applications

Modernizing IT infrastructure
- GPU accelerated compute farms

Big Data, Databases and Messaging applications
- Hadoop, Spark, NoSQL and single instance databases, enterprise business applications

> **Redhat working with Weka and Penguin Computing provides best of breed solutions for AI/ML customers and with a tailored reference architecture make is easy for customers to embark on their AI/ML journey and focus on solving business problems.**
>
> ***Sherard Griffin,*** *Director of Artificial Intelligence Services at Red Hat*

Red Hat

WEKA
910 E Hamilton Avenue, Suite 430, Campbell, CA 95008 T: 408.335.0085 E: info@weka.io www.weka.io