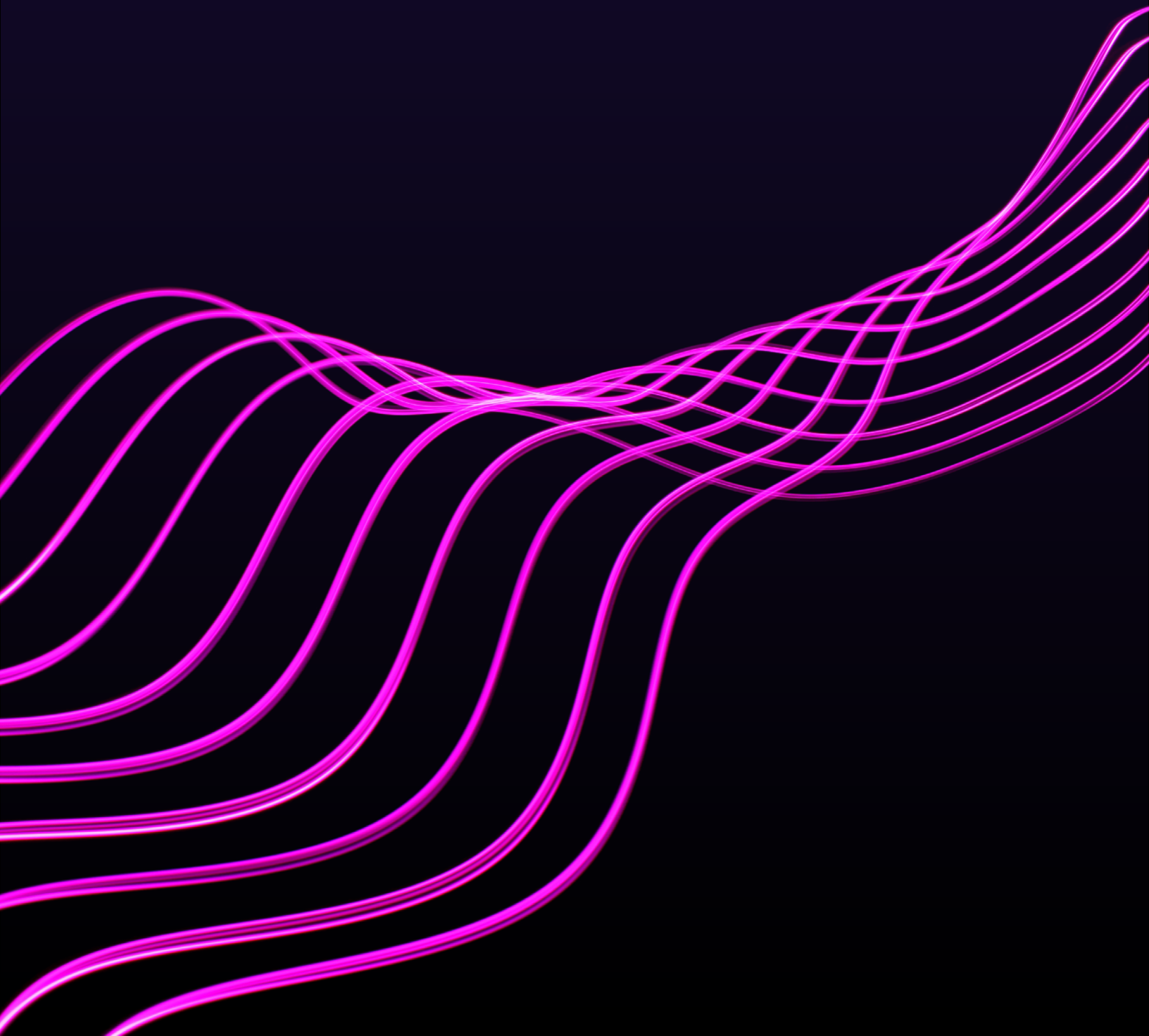
WEKA | run:ai

WEKA and Run:ai

REFERENCE ARCHITECTURE



Contents

Executive

Summary

Run:ai with WEKA is a unique AI architecture that combines innovations from both companies to create an efficient, containerized AI stack that runs both in the private and public cloud. WEKA's software-defined data platform dramatically increases data performance and scale, while Run:ai automates AI workload orchestration and MLOps. Together, these capabilities significantly enhance the efficiency and productivity of your team.

You will be able to innovate faster with less risk. Run experiments and train models across multiple clouds at high speed—with the same proven stack everywhere. Achieve the best return on investment. Fully utilize all GPU and storage resources across multiple clouds to run workloads more efficiently in less time. Start small and scale in a cloud-smart way.

Tested Solution Details

Product Name	Product Version	WEKA Version	Source	K8s Version
Run:ai	2.8	4.01	AWS	1.24

Introduction

Audience

This reference architecture is for those who design, manage, and support WEKA. Consumers of this document should already be familiar with Run:ai and the WEKA platform.

We have organized this document to address critical items for enabling successful design, implementation, and transition to operation.

Purpose

This document covers the following subject areas:

Overview of the WEKA solution.

Overview of Run:ai and its use cases.

The benefits of Run:ai on WEKA.

Recommendations for architecting a complete Run:ai solution on the WEKA Platform, including design and configuration considerations.

Document Version History

Version Number	Published	Notes
1.0	November 2022	Original publication.

Why AI Computing

Artificial Intelligence (AI) is creating new business opportunities for companies in every industry. It is no surprise that Deloitte AI Institute's 2022 State of AI in the Enterprise found that 76% of their survey respondents are planning to increase their expected AI investment in the next fiscal year.

But operationalizing machine learning (ML) and deep learning (DL) requires the ability to process massive amounts of data from different sources in the shortest possible time. As a result, a new wave of accelerated compute (e.g., GPU) devices on the market and rapidly adopted by enterprises. Furthermore, the 10-100x performance over traditional CPUs provided by these GPUs coupled with a massive increase in Deep Neural Network (DNN) models. As a result, it resulted in a Cambrian explosion in AI.

As your organization expands the scope and scale of its AI efforts, success hinges on the ability to deploy complete AI stacks easily—both on-premises and in the cloud—while ensuring that you deliver:

- Optimized application and data performance and management
- The ability to move data easily between locations and Cloud providers
- Optimized and tested with AI and deep learning applications, network architectures, and commercially available GPUs.

But the reality of today's AI can be a long way from this ideal. When companies scale their AI operations, they often struggle with the following challenges:

- Inefficient resource usage. Data scientists and engineers are struggling to effectively use and get access to GPU resources and wasting time waiting for mundane tasks to finish. Additionally, they are unable to move data between locations efficiently.
- Diverse tools. Complexity in dealing with the significant differences in the tools and operational processes across different Cloud providers for containerized AI workloads. These tools make it difficult for teams to take advantage of the resources in each location.
- Complex and slow deployments. Deploying a new AI stack to support experiments or production is complex, time-consuming, and error-prone.
- Data-starved GPUs can have an insatiable appetite for data. But, unfortunately, data bottlenecks mean that GPUs often sit idle because data sets aren't ready or data platforms can't keep delivering data fast enough. These delays increase costs and delay results.

WEKA has joined forces with Run:ai to help you address these challenges, increasing the success of your machine learning and deep learning efforts. Research teams can gain on-demand access to resources for their entire AI workflow, from building the model to training to inference.

WEKA

WEKA Data Platform is deployed on commercially available NVMe servers. An entry-starting cluster size requires eight server nodes for full availability with the ability to survive up to a two-node failure. Each server has a CPU, NVMe storage, and high-bandwidth networking. The exact configuration for the reference architecture is detailed in the Technology Requirements section. The WEKA Data Platform cluster can quickly scale to thousands of nodes.

Performance at Scale

WEKA Data Platform is the world's fastest and most scalable POSIX-compliant parallel file system. WEKA Data Platform is designed to transcend the limitations of legacy file systems that leverage local storage, NFS, or block storage, making it ideal for data-intensive AI and HPC workloads. WEKA Data Platform is a clean sheet design integrating NVMe-based flash storage for the performance tier to the GPU servers, object storage, and ultra-low latency interconnect fabrics such as 200GbE or InfiniBand into an NVMe-over-Fabrics architecture. This creates a highly high-performance scale-out storage system. WEKA Data Platform performance scales linearly as more servers are added to the storage cluster allowing the infrastructure to scale with the increasing demands of the business.

Multi-Protocol Ready

In addition to POSIX access, WEKA Data Platform supports all the standard file access protocols, including NFS, SMB, and S3, for maximum compatibility and interoperability. Hadoop and Spark environments also benefit from the performance of a shared file system through a fully integrated connector that allows WEKA Data Platform to replace HDFS and function as a single, easy-to-manage data lake for all forms of analytics.

Expandable Global Namespace over S3 Object Store

WEKA Data Platform delivers best-of-breed performance from the NVMe flash tier. The namespace can expand to any S3 object store, on-premises, or in the cloud. This optional hybrid storage model with the ability to develop the global namespace to lower-cost hard disk drives in an object store delivers a cost-effective data lake without compromising performance. The integrated tiering to multiple S3 targets enables cost-effective data lifecycle management for older or less used training data.

Advanced Durability and Security

Large and agile datasets in AI/ML frequently require a data versioning capability. This is achieved using WEKA's instant and space-efficient snapshots capability for experiment reproducibility and explainability. The snap-to-object feature captures a point-in-time copy of the entire, unified [flash and object store] file namespace that can be presented as another file namespace instance in a private or public cloud. With crucial management integration, WEKA's integrated snapshots and end-to-end encryption features ensure that data is always backed up and secure throughout its lifecycle. WEKA Data Platform also provides immutability and data mobility for these datasets with instant recovery. WEKA Data Platform can seamlessly back up to multiple cloud targets providing backup, DR, and data governance capability.

Cloud Bursting and Data Mobility

In addition to delivering versioning, WEKA's snap-to-object feature offers additional benefits beyond backup and DR to the public cloud; it enables secure data portability from on-premises to the public cloud for organizations that require access to on-demand GPU resources in the public cloud.

Container Support

Organizations are increasingly adopting containers deployed on Kubernetes (K8s) for AI workloads. Using the WEKA Data Platform K8s CSI plug-in, organizations now have flexibility in how and where they deploy containerized applications. It provides easy data mobility from on-premises to the cloud and back while delivering the best storage performance and latency. Figure X provides an overview of WEKA Data Platform in a typical production deployment.

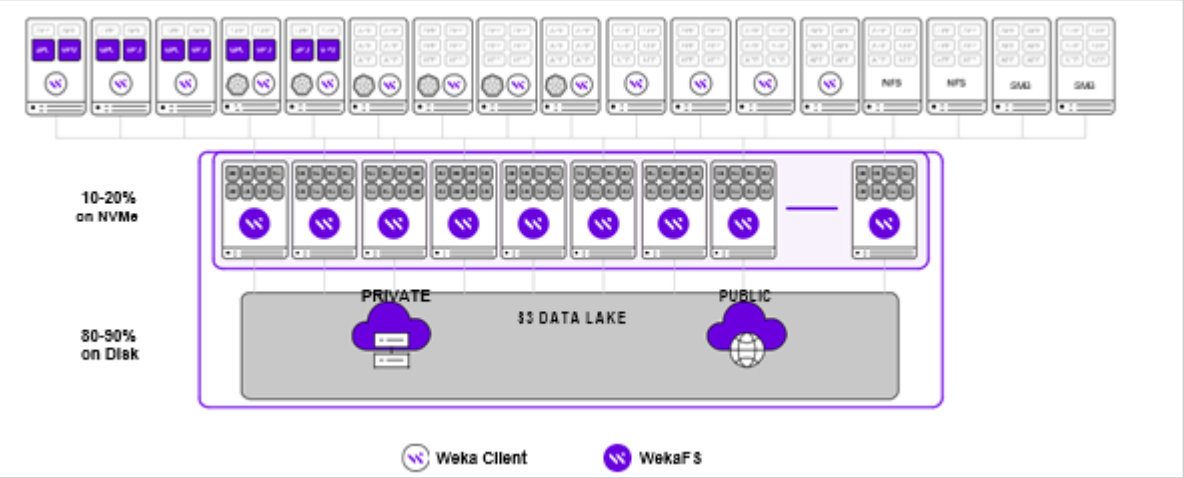
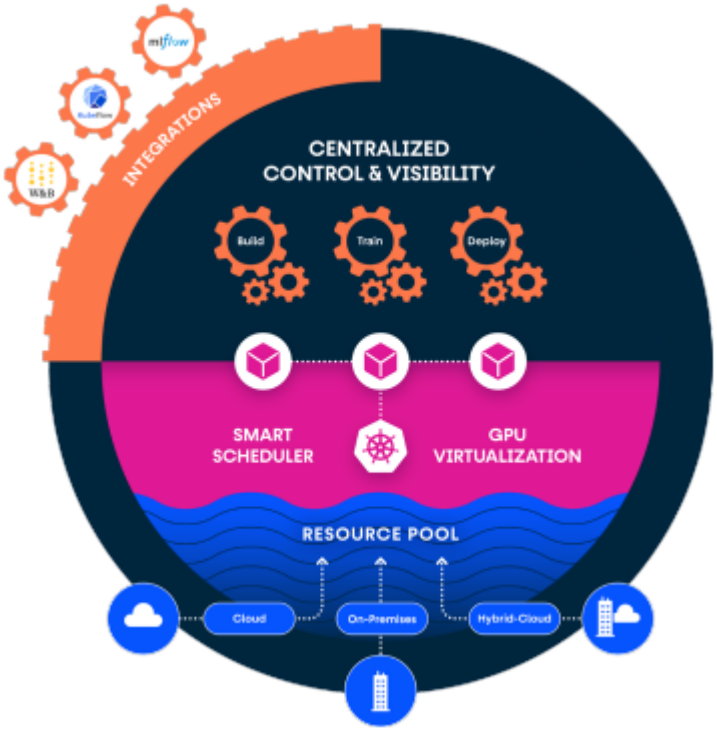


FIG. 1 WEKA Data Platform

Run:ai Architecture

Overview of Run:ai

Run:ai helps organizations simplify and deliver faster on their AI journey from beginning to end. An AI Computing Platform powered by a cloud-native operating system can support running your AI initiatives on-premises, on edge, or in the cloud. The Run: ai Atlas platform gathers the compute and GPU resources in a centralized resource pool and then uses a Kubernetes-based Smart Workload Scheduler to ensure dynamic allocation of resources. Deep integration with GPUs allows the effective sharing of these resources across the entire AI workflow. AI practitioners can easily consume resources in a self-service model using the built-in engines to build, train and deploy or by using 3rd party integrations, such as MLflow, Kubeflow, Weights & Biases, and many more.



Benefits of Run:ai include:

- **Centralize Control and Visibility –** Run:ai Atlas offers dashboards and analytics, giving IT insight across all resources and workloads. Align resource allocation to business goals by setting policies and priorities across departments, projects, or users.
- **Optimize GPU Utilization and ROI –** Automated resource management and efficient sharing of GPU resources enable organizations to achieve higher utilization and increase value per GPU.
- **Build on a Truly Open and Extendable Platform –** Use the built-in workflows optimized for the complete AI development lifecycle or extend the platform by seamlessly integrating with any AI and ML applications, including MLOps tools.

- **Accelerate Hybrid Cloud –** Run:ai has the unique capability to deliver centralized control and visibility across resources on-premises or in the cloud, enabling organizations to make hybrid cloud AI infrastructure a reality.

Using Run:ai with WEKA

Run:ai and WEKA create a unique AI architecture that combines innovations to develop an efficient, containerized AI stack that runs both on-premises and in the cloud. WEKA's software-defined data platform dramatically increases data performance and scale, while Run:ai automates AI workload orchestration and MLOps. Together, these capabilities significantly enhance the efficiency and productivity of your team.

Run:ai and WEKA is the AI/ML and GPU-heavy HPTC workflows platform enabling your team to

- Innovate faster with less risk. Run experiments and train models across multiple clouds at high speed—with the same proven stack everywhere.
- Achieve the best return on investment. Fully utilize all GPU and storage resources across multiple clouds to run workloads more efficiently in less time.
- Start small and scale in a cloud-smart way. Invest in your AI/ML workflows and less on infrastructure using a hybrid multi-cloud approach to grow and scale.

Why Run Run:ai with WEKA

This validated design combines innovative WEKA and Run:ai technology with a best-of-breed data center or public cloud components to create a complete and fully functional AI stack that reduces the burdens on AI infrastructure teams and researchers.

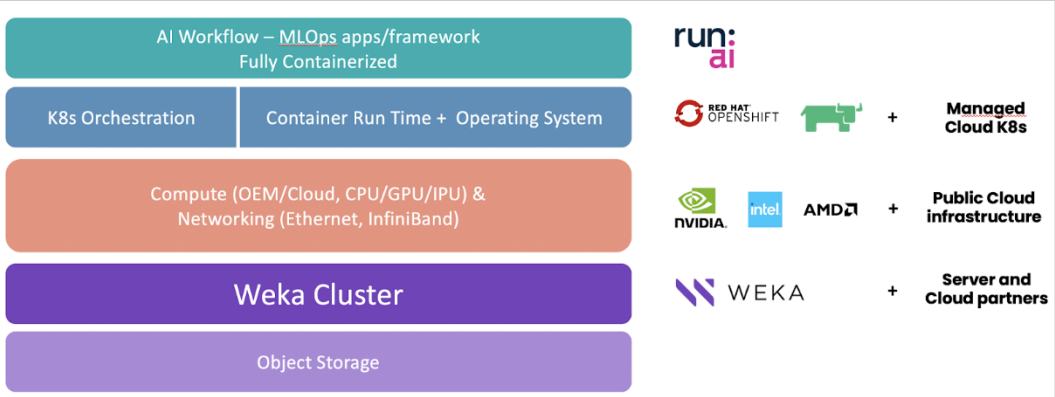


FIG. 2 Run:ai and WEKA Data Platform Stack make AI Compute Easy

At the top of the stack, Run:ai ensures that the control plane for initiating and controlling your AI workflows is always the same whether the stack runs on-premises or in the cloud.

WEKA serves as the stack's foundation, ensuring that all workflows take advantage of WEKA's simplicity, speed, and scale. WEKA can deliver bandwidth that scales to hundreds of gigabytes per second with latency below 250 microseconds.

Kubernetes orchestration is internal to the stack, ensuring workloads are containerized for efficient execution, with open standards for guaranteed portability. Data center deployments utilize K8s distributions from our proven partners, while managed K8s cloud services simplify management for cloud deployments.

For compute and networking, Run:ai and WEKA utilizes best-of-breed hardware to accelerate operations, with the ability to use the most powerful CPUs, GPUs, and IPUs. In the cloud, Smart Stack can take full advantage of the underlying capabilities available in the cloud(s) you choose.

Solution Design

Design Decisions

The following tables cover design decisions and rationale for Run:ai with WEKA

Table 1: General Design Decisions: WEKA

Item	Details	Rationale
Minimum Size	6 WEKA Servers 2 Compute+GPU Servers	Minimum size requirements
Scale approach	Incremental, modular scale	Allow for growth from PoC to a massive scale
Scale Units	Servers	Granular scale to precisely meets the capacity demands.

Table 2: General Design Decisions: Networking

Item	Details	Rationale
Compute/K8s	25GB or greater	The network between K8s control plain and workers
Storage	25GB or greater	WEKA can utilize higher network speeds

Run:ai Sizing

The section will include the requirements that are specific to Run:ai

Hardware (K8s Control Plane)

Item	Details	Rationale
CPU	4 CPU	Minimum CPU
Memory	8 GB	Minimum Memory
Storage	120 GB	Minimum Storage
Nodes	1	Number of Master Nodes (Minimum)

Hardware (K8s Worker)

Item	Details	Rationale
CPU	4 CPU	Minimum CPU
Memory	8 GB	Minimum Memory
Storage	120 GB	Minimum Storage
Worker Nodes	2 Nodes, 1 Node must have a GPU (NVIDIA)	Number of Worker Nodes (Minimum)

Hardware (K8s Worker with GPU)

Item	Details	Rationale
CPU	4 CPU	Minimum CPU
Memory	8 GB	Minimum Memory
Storage	120 GB	Minimum Storage
GPU	https://docs.nvidia.com/datacenter/cloud-native/gpu-operator/platform-support.html#supported-nvidia-gpus-systems	A typical use case could see 2-8 Jobs running on the same GPU
Worker Nodes	2 Nodes, 1 Node must have a GPU (NVIDIA)	Number of Worker Nodes (Minimum)

Software

Item	Details	Rationale
Kubernetes versions	1.21 through 1.24 (1.25 is not supported)	K8s requirements
Kubernetes providers	RedHat OpenShift, Rancher KE, HPE Ezmeral, CNCF Certified Kerbernetes environment	Tested Providers (OCP 4.8 to 4.10)
Kubernetes cloud providers	EKS, AKS	Tested Cloud Providers
NVIDIA GPU Operator	1.9 or higher	Required version
Linux version	Ubuntu 20.04 and CoreOS for OpenShift (Tested)	Tested versions
Shared Data volumes	NFS and K8s CSI (WEKA POSIX Client)	Current supported Storage

Recommended order of Deployment

1. Install WEKA in AWS
2. Create Kubernetes cluster in ECS (Control Plane + Worker + Worker with GPU)
3. Mount all servers to the WEKA filesystem
4. Create the CSI connection

Burst to Cloud

The WEKA Snap-To-Object solution can be used to simplify data management at scale for customers facing data mobility and data protection challenges for large data sets. This feature enables committing a full copy of the snapshot of the data to an object store, which can be used to restore the data on the original WEKA cluster or onto another WEKA cluster. After the first snapshot-to-object has been completed, subsequent snapshots are stored incrementally, so commit time is limited to the changes and extremely fast. WEKA Data Platform also supports sending snapshots to a remote object storage system in parallel using the Remote Backup feature. This leverages the incremental nature of snapshots by only sending the changes across the wire to the destination object-store.

Setup

- Run:ai and WEKA CSI setup and operation
- Submit a job into production, if more than 10 jobs are running locally (This can be adjusted)
- Snap to Object
  - Create a snapshot
  - Sync the snap to the secondary site
  - Burst - bring the data online on the secondary site
- Automatically Schedule the job on the burst cluster

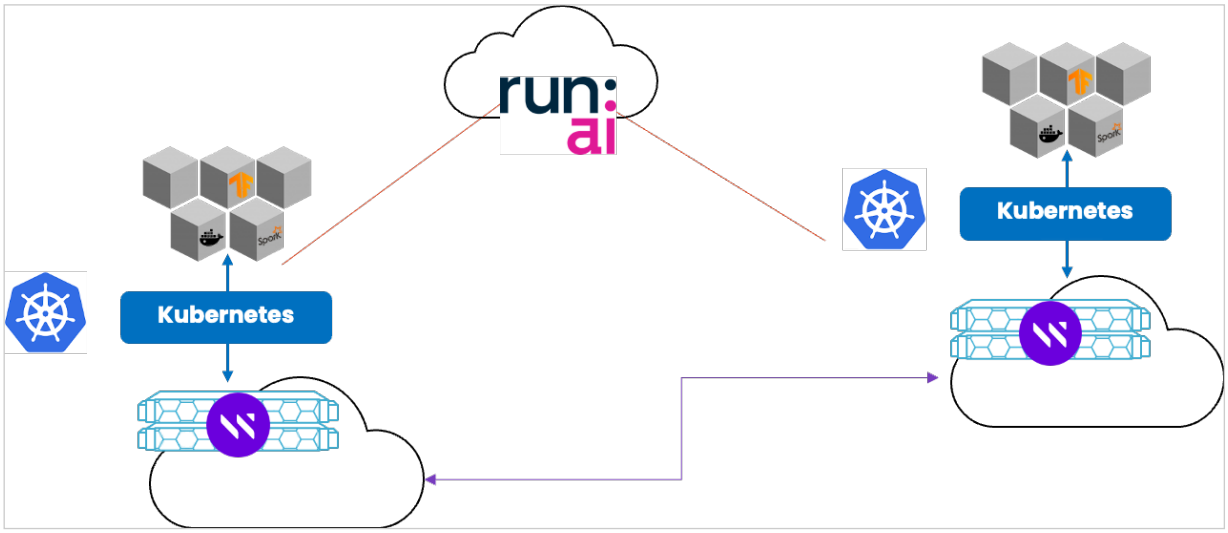


FIG. 3 Run:ai and WEKA allow organizations to create a hybrid cloud AI Compute Platform

Best Practices for Run:ai with WEKA

Network Time Protocol (NTP)

Use the Network Time Protocol (NTP) to synchronize the clocks on all servers.

Resiliency

The WEKA system is a distributed cluster protected from 2 or 4 failure domain failures, providing fast rebuild times as described in the WEKA system overview section.

Instance failure

If an instance failure occurs, the WEKA system rebuilds the data. Add a new instance to the cluster to regain the reduced compute and storage due to the instance failure.

Upload snapshots to S3

Run:ai and WEKA advise using periodic (incremental) snapshots to back up the data and protect it from multiple EC2 instances failures.

The recovery point objective (RPO) is determined by the cadence in which the snapshots are taken and uploaded to S3. The RPO changes between the type of data, regulations, and company policies, but it is advisable to upload at least daily snapshots (Snap-To-Object) of the critical filesystems.

If a failure occurs and it is required to recover from a backup, spin up a cluster using the Self-Service Portal or CloudFormation, and create filesystems from those snapshots. You do not need to wait for the data to reach the EC2 volumes. It is instantly accessible through S3. The recovery time objective (RTO) for this operation mainly depends on the time it takes to deploy the CloudFormation stack and is typically below 30 min.

Snap to Object Best Practices

The following are the best practices for WEKA Snap to Object.

- The number of buckets required for the public and private clusters are the same
- User/System configurations (WEKA Data Platform, Protocols, Mgmt, Quotas) are not synced. (The S3 Configuration on site A not replicated to site B)
- WEKA Data Platform on the target site has to be 100GB
- Obtain the right object locator for the snapshot to be restored
- For incremental snapshot download feature, you are required to be in release 4.x and above
- Close all open files in Site B before file system refresh operation
- Local changes made to the synchronized file system on Site B will be lost. To preserve changes, copy them to a different file system in Site B

Dockers, Images, and Kubernetes

Researchers are typically proficient in working with Docker. Docker is an isolation level above the operating system, which allows creating your bundle of the operating system + deep learning environment and packaging it within a single file. The file is called a **docker image**.

You create a **container** by starting a docker image on a machine.

Run:ai is based on **Kubernetes**. At its core, Kubernetes is an orchestration software above Docker: Among other things, it allows location abstraction as to where the actual container is running. This calls for some adaptation to the Researcher's workflow as follows.

Image Repository

If your Kubernetes cluster contains a single GPU node (machine), then your image can reside on the node itself (in which case, when Run:ai submit workloads, the Researcher must use the `flag --local-image`).

If your Kubernetes cluster contains more than a single node, then, to enable location abstraction, the image can no longer reside on the node itself. It must be relocated to an image repository. There are quite a few repository-as-a-service, most notably Docker hub. Alternatively, the organization can install a private repository on-prem.

Day-to-day work with the image located remotely is almost identical to local work. The image name now contains its location. For example, `nvcr.io/nvidia/pytorch:19.12-py_3` is a PyTorch image that is in **nvcr.io**. This is the Nvidia image repository as found on the web.

Data

Deep learning is about data. It can be your code, the training data, saved checkpoints, etc. If your Kubernetes cluster contains more than a single node, then, to enable location abstraction, the data must sit outside the machine, typically on network storage. The storage must be uniformly mapped to your container when it starts (using the `-v` command).

Working with Containers

Starting a container using docker usually involves a single command line with multiple flags. A typical example:

```
docker run --runtime=nvidia --shm-size 16G -it --rm -e HOSTNAME='hostname' \

      -v /raid/public/my_datasets:/root/dataset:ro -i nvcr.io/nvidia/pytorch:19.12-py3
```

The docker command `docker run` should be replaced with a Run:ai command `runai submit`. The flags are usually the same, but some adaptation is required. A complete list of flags can be found here: Run:ai submit.

There are similar commands to get a shell into the container (`runai bash`), get the container logs (`runai logs`), and more. For a complete list, see the Run:ai CLI reference.

Schedule an Onboarding Session

It is highly recommended to schedule an onboarding session for Researchers with a Run:ai customer success professional. Run:ai can help with the above transition, but adding to that, we at Run:ai have also acquired a large body of knowledge on data science best practices which can help streamline Researchers' work as well as save money for the organization.

Conclusion

With Run:ai and WEKA, your teams can deploy new AI stacks whenever needed—including additional cloud stacks for short-term bursting or long-term needs—in less time and with less hassle.

Regardless of location, each new AI stack operates identically, with the same control plane, tools, and capabilities. Your team can start using the new stack without learning about any new tools, limitations, or other peculiarities.

No matter what underlying infrastructure you choose for the new stack, you can also be confident that Run:ai and WEKA will deliver the best possible performance from the resources you give it, ensuring you get the maximum return on your investment in cloud infrastructure or additional on-prem hardware.

Appendix

References

Product & Technology	
WEKA File System	
WEKA Data Platform Datasheet	https://www.weka.io/wp-content/uploads/files/2022/10/about-weka-data-platform.pdf
WEKA Data Platform Architecture White Paper	https://www.weka.io/wp-content/uploads/files/2017/12/Architectural_WhitePaper-W02R6WP201812-1.pdf
Run:ai	
Run:ai Solution Brief	https://pages.run.ai/hubfs/PDFs/RunAI-Solutions-Brief-Dec-2020.pdf

About WEKA

WEKA offers a modern subscription software-based data platform delivering 10x+ performance and scale demanded by today's cloud and AI workloads. With the simplicity of NAS, the performance of SAN or DAS, and the scale of object storage, no more compromises between Simplicity, Speed, or Scale. Learn more at www.weka.io or follow us on Twitter @WEKAIO.

WKA339-01 01/2023