



451 Research Market Insight Report Reprint

WEKA hones focus on emerging sustainable AI and GPU cloud opportunities

February 13, 2024

by **Henry Baltazar**

The vendor is positioning its offering as a premier data infrastructure platform for AI- and performance-intensive workloads. Its parallel storage technology is effective both on-premises and in public cloud environments, which makes it a viable option for hybrid and multicloud infrastructures.

S&P Global
Market Intelligence

This report, licensed to WEKA, developed and as provided by S&P Global Market Intelligence (S&P), was published as part of S&P's syndicated market insight subscription service. It shall be owned in its entirety by S&P. This report is solely intended for use by the recipient and may not be reproduced or re-posted, in whole or in part, by the recipient without express permission from S&P.

Introduction

WEKA is positioning itself as a provider of a premier data infrastructure platform for AI- and performance-intensive workloads. The vendor's data platform leverages high-performance parallel file system technology at its core and is effective across both on-premises and in public cloud environments, which makes it a viable option for hybrid and multicloud infrastructures. The emergence of graphics processing unit (GPU) clouds has opened new opportunities for WEKA as service providers look to cash in on the generative AI market opportunity. The rising importance of ESG (environmental, social and governance) has highlighted the need for sustainable AI architectures, which the company's Data Platform can facilitate.

THE TAKE

Although discussions around artificial intelligence infrastructure often focus on GPUs, data platforms provide key capabilities, including data management, storage and migration. The performance of storage also has a direct impact on the rest of the infrastructure, since GPU performance cannot be optimized if the processors are waiting for a slower storage infrastructure to feed them data. WEKA's latest announcements highlight its advances in performance-sensitive use cases and its ability to manage and migrate data within modern hybrid cloud and edge environments.

In the public cloud space, the vendor's Converged Mode for Cloud on AWS offering has emerged as a cost-efficient alternative to Amazon FSx for Lustre, which it could extend to other public clouds to create similar opportunities for public cloud HPC and AI use cases. The rise of GPU clouds to support AI/machine-learning and generative AI workloads has also created partnership opportunities with Applied Digital Corp., Iris Energy Ltd. and WEKA's newest partner, NexGen Cloud. Intense competition in the storage for AI space should highlight WEKA's ability to work with a wide range of customers and service providers in various verticals.

Context

WEKA (formerly known as WekaIO) was founded in 2013 and is headquartered in Campbell, Calif., with research and development sites in Israel, India, Ukraine and the US. The vendor has over 250 customers, including 11 members of the Fortune 50, and closed out 2023 with 100% year-over-year growth. It has a headcount of 350 and has raised over \$275 million in funding. WEKA's last major round was a series D that closed in November 2022 to the tune of \$135 million and was led by Generation Investment Management, with participation from new and existing investors 10D, Atreides Management, Celesta Capital, Gemini Israel Funds, Hewlett Packard Enterprise Co., Hitachi Ventures, Key1 Capital, Lumir Ventures, Micron Ventures, Mirae Asset Capital, MoreTech Ventures, Norwest Venture Partners, NVIDIA Corp., Qualcomm Ventures and Samsung Catalyst Fund.

In September 2023, the company was named the official technology partner of rock band U2, providing the critical data infrastructure needed to support its U2:UV Achtung Baby Live residency at Sphere in Las Vegas. During the show, the WEKA Data Platform supports the streaming of 402 GB of 16,000 resolution video content per second on Sphere's four 160,000-square-foot LED displays along with audio to over 164,000 independent channels — it also helped migrate archival video footage rendered in the UK to on-premises servers in Las Vegas via the cloud. The show's video archive initially started at 500 TB but quickly tripled in size to 1.5 PB as the band continued to add new content to enhance the live event experience.

Sustainable AI initiative

Generation Investment Management's investment in WEKA is notable since it was part of the firm's Sustainable Solutions Fund IV, which is focused on deep research into "green data" companies that can help address the growing energy consumption problem associated with the exponential expansion of workloads such as AI/ML HPC. As such, WEKA's strategy has evolved to emphasize the efficiency and sustainability benefits of its storage platform, which is becoming more relevant with the increasing influence of ESG requirements. In 451 Research's Voice of the Enterprise: Storage, ESG Attitudes 2023 survey, 56% of respondents noted that power efficiency and rack-space consumption were very important when evaluating which storage platforms should be installed in their datacenters.

WEKA claims that its technology can help customers reduce over 260 tons of CO₂e (carbon dioxide equivalent) per petabyte over the typical three- to five-year storage equipment life cycle compared with traditional storage architectures. On a storage level, the reduction the vendor offers comes from its data copy reduction and cloud elasticity capabilities, which it claims will lower a customer's data infrastructure footprint four to sevenfold.

Beyond storage, the WEKA Data Platform can also greatly enhance the data-processing and power consumption efficiency of GPUs by ensuring that they are constantly fed data. Given that idle servers still draw 50% of the power as active ones, by keeping GPUs occupied an organization can greatly minimize the amount of wasted energy consumed in their environments.

Partners accelerate expansion into GPU and public clouds

In the public cloud space, WEKA's Converged Mode for Cloud on AWS is a cloud storage service that taps into the unused memory and flash storage resources within AWS high-performance instances. Open-source generative AI service provider Stability Ai was previously utilizing Amazon FSx for Lustre to ensure that its storage stack had enough performance to feed hundreds of Amazon EC2 P5 instances running its training and inference of AI models in the cloud. Through its partnership with WEKA, Stability Ai is now deploying the WEKA Data Platform on its Amazon EC2 P5 instances and using the flash storage and memory included in those instances to replace the Amazon FSx for Lustre storage service. Converged Mode for Cloud was released as a public preview for AWS deployments in 2023, with additional cloud support coming in the future.

In October 2023, WEKA announced a partnership with Iris Energy, which is the owner and operator of institutional-grade bitcoin mining datacenters that are powered by 100% renewable energy. In the partnership, WEKA offers storage and data management to support Iris' new GPU cloud services for generative AI and other performance-intensive workloads.

In November 2023, the vendor announced a partnership with Applied Digital that provides next-generation digital infrastructure for HPC workloads. Applied Digital's new AI Cloud Service uses the WEKA Data Platform to optimize its data pipelines and boost GPU datacenter efficiency, and WEKA's storage provides 1.8-TBps bandwidth performance from a single rack of equipment. Beyond the infrastructure and energy cost savings, the companies claim that WEKA's Data Platform has allowed them to accelerate GPU utilization by up to 20x while reducing training time by 10x-100x.

Additionally, WEKA and partner NexGen Cloud just announced that WEKA would be providing the high-performance storage infrastructure for the latter's \$1 billion AI Supercloud, which was created with the goal of offering sustainable and affordable infrastructure to democratize AI. After evaluating multiple storage tools, NexGen Cloud selected WEKA because its hardware-agnostic software allowed it to leverage existing hardware to enhance the efficiency and sustainability of its cloud offerings, in contrast with other storage suppliers that pitched hardware appliances.

Competition

In the unstructured data storage space, Dell Technologies Inc., NetApp Inc., HPE, Hitachi Vantara and IBM Corp. have been industry leaders for decades now, and they all have deep product portfolios that cover file and object storage with their own products or via partnerships. WEKA has an OEM partnership with Hitachi Vantara to provide high-performance unstructured storage and data management, which complements its existing storage platforms. Several other players have become significant, including VAST Data, Pure Storage Inc., Hammerspace, DDN, Qumulo and Scality, all of which are also pursuing the generative AI storage opportunity.

In the public cloud segment, the need for scalable, high-performance storage has driven the major players to add parallel file system offerings, such as the new Azure Managed Lustre service and Amazon FSx for Lustre, to their storage portfolios to help customers keep up with demanding AI- and HPC-class workloads. WEKA's software has already displaced these cloud services by providing performance and cost advantages over public cloud offerings, which should broaden the appeal of its technology for cloud stakeholders that do not have familiarity with enterprise storage products.

SWOT Analysis

STRENGTHS WEKA's technology provides the high performance and data accessibility required for modern workloads. The vendor has had success in a number of verticals, including financial services, life sciences, and media and entertainment.	WEAKNESSES The company is still relatively small compared with the top storage players, and will need to boost awareness of its data platform.
OPPORTUNITIES AI and other use cases that require the management of large-scale unstructured data are increasing rapidly and creating new opportunities such as GPU cloud.	THREATS Every on-premises infrastructure and public cloud provider is aggressively targeting the same AI storage infrastructure opportunities.

CONTACTS

The Americas

+1 877 863 1306

market.intelligence@spglobal.com

Europe, Middle East & Africa

+44 20 7176 1234

market.intelligence@spglobal.com

Asia-Pacific

+852 2533 3565

market.intelligence@spglobal.com

www.spglobal.com/marketintelligence

Copyright © 2024 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON “AS IS” BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT’S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence’s opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global Market Intelligence does not endorse companies, technologies, products, services, or solutions.

S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global’s public ratings and analyses are made available on its websites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.