WEKA

SOLUTION BRIEF

Save 260 tons of CO_2e per PB with the WEKA Data Platform

Challenges

- Data centers consume more than 3% of global energy consumption which is projected to rise to 8% by 2030 if left unchecked
- Legacy data architectures have a greater environmental impact than contemporary, modern approaches, which can negate sustainability efforts

Solution

- The WEKA® Data Platform drives 10x-50x better AI/ML stack efficiency reducing annual GPU operating energy
- WEKA also lowers the data infrastructure footprint by 4x-7x through data copy reduction and cloud elasticity.

Benefits

- Reduced energy consumption while also delivering faster results
- Over 260 tons of CO2e per petabyte saved compared to a traditional data architecture

The increased pressure on organizations to deliver data-driven insights and business outcomes has created an exponential demand for power in modern data centers, both on-premises and in the cloud, making them some of the world's biggest consumers of power. Evidence suggests that today, data centers account for roughly 3%[1] of global energy consumption; left unchecked, that is projected to rise to 8% by 2030.[1]

Organizations are implementing energy reduction, space consolidation, and green utility grids to mitigate this growth. However, there are still significant hurdles to data center sustainability efforts, and further improvements will be necessary to limit energy and emissions as new needs arise. The demand for high-performance computing continues to grow, requiring massive data infrastructure to support new data-intensive artificial intelligence (AI) and machine learning (ML) workloads and applications. As organizations respond to these demands, they are challenged by legacy data architectures that have a greater environmental impact than contemporary, modern approaches, which can negate their sustainability efforts. Siloed applications, excessive data movement, and the need to oversize an environment to meet performance goals all lead to greater energy consumption and, as a result, more carbon emissions.

Performance-Intensive Workloads Create Energy Waste

Data-driven innovation through modern workloads like AI, ML, and high-performance computing (HPC) is driving a move from periodic and slower batch processes to continuous high-speed data pipelines. These GPU-accelerated, data-intensive workloads consume data significantly faster

1 https://dl.acm.org/doi/pdf/10.1145/3483410

KEY INSIGHT

What Is CO_2e and Why Is It an Important Measure?

While CO_2 refers to Carbon Dioxide, a natural and odorless greenhouse gas resulting from the burning of fossil fuels, CO_2e, or Carbon Dioxide Equivalent, is representative of the collective greenhouse gasses. CO_2, the most prevalent greenhouse gas after water vapor, has become the proxy by which we measure greenhouse gas emissions. This measure has been developed by scientists to express emissions in terms of their relative global warming potential (GWP), where CO_2 has a GWP of 1. In this way, we can express greenhouse gas emissions as an equivalent of CO_2 using the GWP principle.[3]

than CPU-based ones. Traditional data infrastructure, which was designed for legacy workloads, can't keep up with today's data pipelines creating bottlenecks that leave GPUs waiting for data up to 70% for the time[2] and idling servers draw up to 50% of maximum power, wasting energy. The faster data can reach the cores, the more quickly it can be processed to reach the outcomes your organization needs and thus reduce energy waste.

But here's the reality: each one of the steps in a data pipeline usually has a completely different profile based on the data type. For example, some of these steps have different I/O demands as data is processed, such as the need for low latency, perhaps random small I/O. In contrast, others need massive streaming throughput or a concurrent mix of the two because of sub-steps within the process(es). In most environments, multiple pipelines will run concurrently but at different stages, amplifying the need to handle different I/O profiles simultaneously.

The variability of these demands can cause issues with legacy storage, which is inherently tuned to address either one data type or a specific performance profile. As a result, the solution is to execute a copy strategy across multiple siloes of the legacy storage tuned for the right profile. This additional management complexity and multi-copy redundancy lead to excess storage infrastructure, slower access cycles, and increased energy consumption with an associated carbon footprint.

Get Better Performance *and* Environmental Benefits

The WEKA® Data Platform is a software-defined solution that is purpose-built for the most performance-intensive applications and large-scale AI, ML and HPC workloads running on-premises and in multicloud environments. Its advanced, cloud-native architecture sustainably delivers speed, simplicity, and scale, while supporting seamless data-sharing, so you can take full advantage of your enterprise data in virtually any location.

2 https://www.enterpriseai.news/2020/03/17/how-to-beat-the-gpu-storage-bottleneck-for-ai-and-ml/
3 https://www.sustainablebusinesstoolkit.com/difference-between-co2-and-co2e/

KEY INSIGHT

Siloed applications, extensive data movement, and the need to oversize an environment to meet performance goals all lead to greater energy consumption and as a result, more carbon emissions.

The WEKA® Data Platform removes barriers to data-driven innovation through its advanced, high-speed software architecture, which is optimized to solve complex data challenges, drive improved AI/ML and HPC stack efficiencies, and reduce the overall data infrastructure footprint.

The WEKA Data Platform makes all data run as fast as local regardless of its location and increases utilization of on-premises and cloud GPUs. Its Zero Copy architecture also reduces the complexities of having to maintain and copy data between disparate storage systems which delivers a unique strategic advantage to WEKA customers and one that has significant carbon savings and other environmental benefits.

WEKA delivers 10-50x increased GPU stack efficiency through its high-speed data architecture reducing annual GPU operating energy. It also can shrink data infrastructure footprints by 4-7x through data copy reduction and cloud elasticity. Reducing footprint also lowers manufacturing lifecycle impact and data center space requirements, which often exceed the impact of the infrastructure footprint.

Calculating Savings with WEKA

ESG Capital Group conducted a comparative life cycle assessment of the WEKA Data Platform compared to legacy data architectures to determine its potential carbon savings and other environmental benefits.

Legacy data storage architecture was evaluated on the basis of an equivalent amount of data (e.g. petabytes, PB) compared to that managed by WEKA over the course of a year. In addition to the benefits described above, other key factors in the comparison were reduction in server idle time (idle servers draw up to 50% of maximum power), reduction in electricity and overhead such as cooling through efficient storage and computing processes. Also calculated was the impact on carbon-intensive supply chains and short lifespan of legacy storage. The ability to move some workloads to the cloud was also a factor as renewable energy at large cloud datacenters reduces emissions intensity of energy.

Taking into account US and Global research studies to estimate average energy consumption and CO2 emissions generated for equipment and operations, they concluded that the WEKA Data Platform saves over 260 tons of CO2e per petabyte over the typical 3-5 year life cycle compared to traditional data architectures. And these savings will grow with your infrastructure.

Annual carbon emissions (metric tons CO_2e)	**Legacy Infrastructure**	**WEKA**	**Savings**
Hardware CO_2 footprint (per PB)	152 tons CO_2e	30 tons CO_2e	122 tons CO_2e
Operating CO_2 footprint (per PB)	164 tons CO_2e	26 tons CO_2e	138 tons CO_2e
Total CO_2e (per PB)	**316 tons CO_2e**	**56 tons CO_2e**	**260 tons CO_2e**
Average Customer CO_2e (7PB)	**2,212 tons CO_2e**	**392 tons CO_2e**	**1,820 tons CO_2e**

For a typical WEKA customer with over 7 PB of data, this is a significant carbon savings of 1,820 CO2e over the course of the life cycle.

KEY INSIGHT

On average, each WEKA customer saves over 1,820 tons CO_2e over the course of a typical infrastructure life cycle.

Make a Data Platform Strategy Part of Your Sustainability Plans

New approaches are needed to mitigate the environmental impact of the growing demand for high-performance and data-intensive computing. A data platform strategy is one key way to meet your performance goals while lowering energy consumption and reducing your organization's carbon emissions.

Contact us at info@weka.io for more information about how the WEKA Data Platform can be part of your green data initiatives.

WKA329-01 02/2023