# Product Spotlight
# How To Solve the Unstructured Data Paradox - WekaIO

**Storage Switzerland, LLC**

George Crump | Lead Analyst

There is a capacity and performance paradox to unstructured data that wastes IT budget and resources as IT tries to find the perfect solution. In terms of capacity, the organization has more and more data to store each year, and rarely is data removed. It just keeps growing. Storage admins attempt to manage the associated costs by migrating data to more economical storage as the data ages. Storage performance on the other hand is variable. An application will need to quickly process a data set but once that processing is done, the compute is no longer needed.
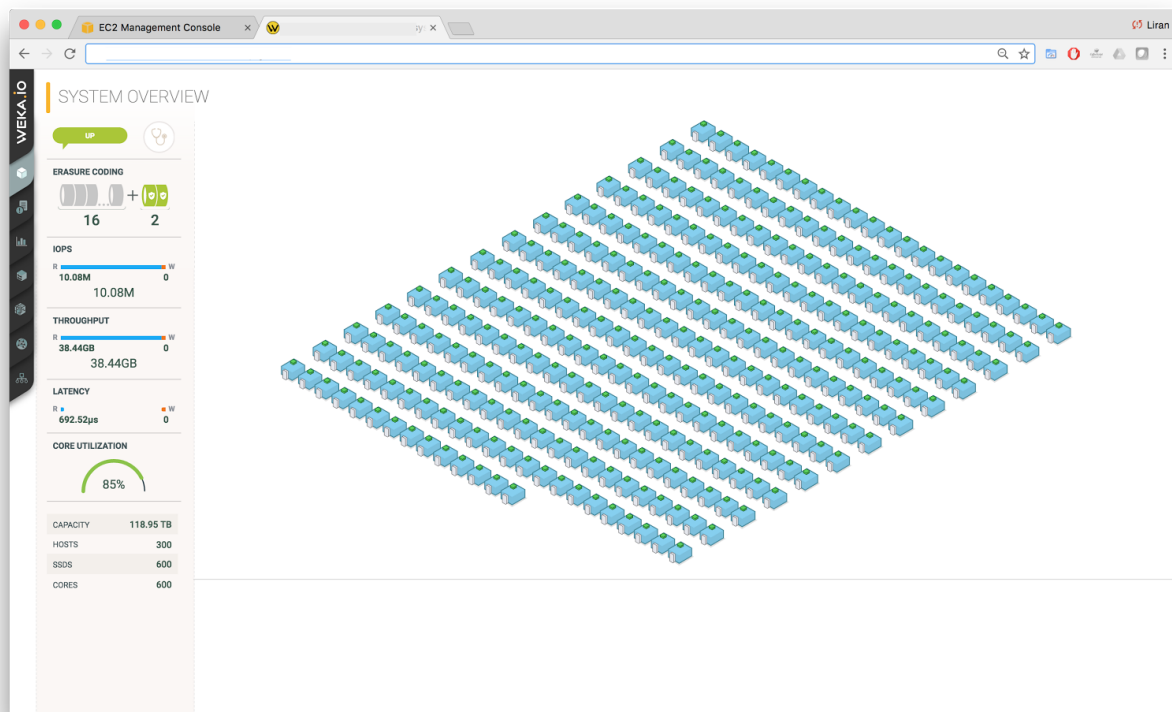
The paradox forces IT planners to choose between scale-up or scale-out file systems and network attached storage (NAS) appliances to deliver high performance access to unstructured data. Both of these architectures, though, have limitations. Scale-up will reach a capacity limit, forcing the purchase of additional systems. Scale-out eventually wastes compute performance because it requires too many nodes to meet capacity expectations. Each of those nodes comes with compute which is not needed, at least most of the time.

## The Unstructured Data Problem

The paradox is caused because unstructured data is unique in how it requires high performance I/O. First, its performance demand is variable. Unlike a typical database requiring high performance with occasional demand for even higher peak performance, unstructured data requires moderate to low performance most of the time but has more frequent peaks, depending on the workload. This leads to unpredictable performance demands and over-provisioning. Second, the workloads themselves are variable. In the database use case, where the workload is consistent, unstructured workloads can vary from processing thousands of small files to sequentially streaming a few big files. Still other applications are latency sensitive.

## Limitations of File-based Solutions

Using NAS architectures as an example, organizations must choose between scale-up or scale-out solutions which may not be flexible enough to adapt to the performance and workload variety. In a scale-up NAS architecture the system is powered by a single NAS controller. The amount of compute it has available for storage performance is fixed as is its capacity. The organization typically has an excess of both initially but runs out of one of the two resources as unstructured data needs scale.

A scale-out architecture has fixed amount of compute and capacity per node. If the organization needs more of either then additional nodes are added. The nodes within the scale-out architecture are clustered creating a single point of management. The most typical reason for expansion is capacity but compute resources come with that addition of capacity, and except when data is processed those resources are wasted, lowering infrastructure efficiency and increasing total cost of ownership.

## WekaIO's Matrix™: A More Flexible, Scalable Architecture

WekaIO's Matrix is a software-only solution designed to bring web-scale performance, flexibility, and simplicity to data centers. It can be deployed as hyperconverged on any existing compute cluster alongside the current workloads, as a dedicated storage server on separate hardware, or entirely in the cloud. In addition, it runs in bare metal, virtual machine, or container environments.

Any node contributing capacity needs to have at least one SSD assigned to it. MatrixFS, the underlying distributed, parallel file system, then aggregates the capacity from all the nodes into a global file system. Data is striped across these nodes with a patented data protection scheme that provides both performance and data resiliency.

Unlike a traditional scale-out architecture, Matrix's compute performance is dynamically tunable. The software fully supports multi-core processors. When a peak workload needs more performance to meet an I/O demand additional cores can be allocated to Matrix to more quickly service the request. After the peak has passed those cores can be released and the compute cores made available to other applications in the cluster.

## Optimizing for Multi-Node Performance

One of the challenges with any scale-out architecture, especially one that is hyperconverged, is the impact of east-west traffic, or inter-node communication. The WekaIO architecture is designed for fast Ethernet and utilizes Single Root I/O Virtualization (SR-IOV) to run storage functions. It also utilizes the Data Plane Development Kit (DPDK) to give its software access to the network without going through the kernel, resulting in significant latency reduction and delivering all-flash performance to networked storage.

Most storage systems have flash as either an option or are all-flash. Few take full advantage of flash instead of continuing to access flash as if it were a legacy hard drive. WekaIO is totally optimized for flash and does away with these legacy techniques.

Finally, flash is still too expensive to store all data all the time, so Matrix has the ability to tier data off to private object storage or a public cloud provider like AWS. But because of Matrix's global file system, access to that data is unchanged. The organization can even leverage the cloud for on-demand burst processing, scaling to thousands of processors to deliver workload analysis.

## StorageSwiss Take

Organizations with performance sensitive or large stores of unstructured data have a new option to consider. In most cases these organizations have a compute cluster already established for processing. WekaIO's proposition is to simplify data storage by making it a scalable application that leverages the compute cluster to provide storage services. It provides very high performance, data resiliency, and cost controls by tiering its data to the cloud. The combination is impressive and certainly worthy of consideration by IT planners.

*Sponsored by Weka.IO*

## WEKA.IO
### Radically Simple Storage™

**About WekaIO**

WekaIO leapfrogs legacy infrastructures and improves IT agility by delivering software-centric data storage solutions that unlock the true promise of the cloud. WekaIO Matrix software is ideally suited for performance intensive workloads such as Web 2.0 application serving, financial modeling, life sciences research, media rendering, Big Data analytics, log management and government or university research. For more information, visit http://www.weka.io

**About Storage Switzerland**

Storage Switzerland is an analyst firm focused on the virtualization and storage marketplace. For more information please visit our website **storageswiss.com**