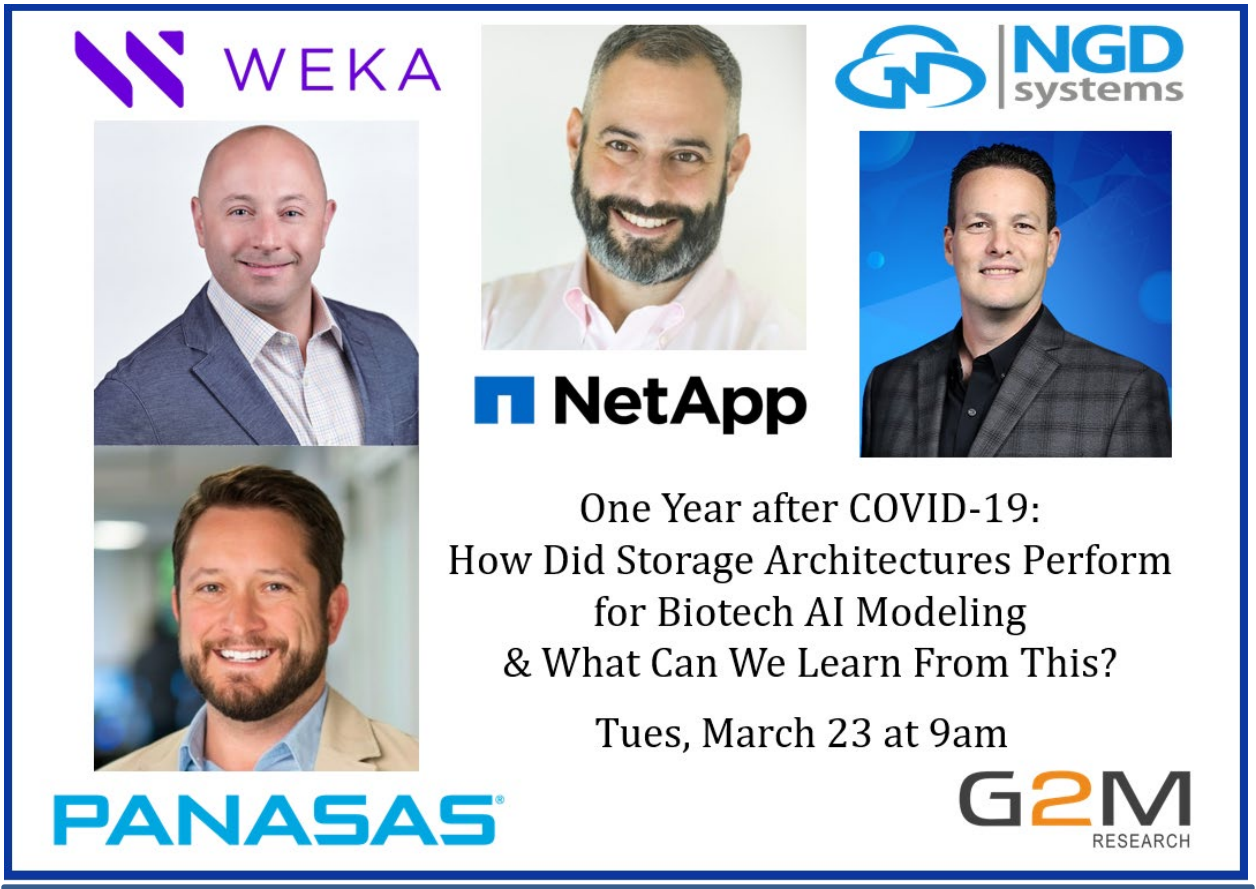
Highlights

Over 900 people registered for our last webinar, with sponsors Weka, Kioxia, and NVIDIA. The advantage of registering, even if you cannot attend, is the webinar link and materials are emailed directly to registrants. We can appreciate that most people are too busy to attend webinars during the workday but still want look through the slide deck, particularly for webinars featuring major enterprise storage companies, startups, new product solutions, and providing meaningful approaches to data management and storage challenges.

If you are interested in sponsoring a webinar but don't see a topic that quite fits, we can modify topics or add topics to meet your objectives.

Cheers! Mike Heumann



Moving AI from the Cloud and onto the Edge

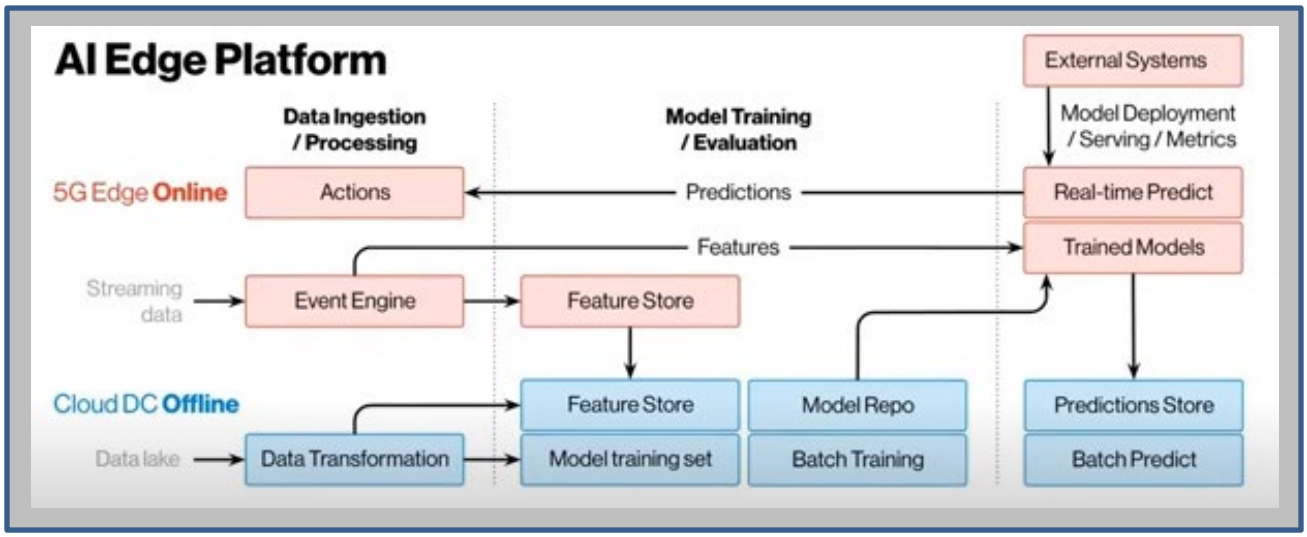
verizon media

Historically, artificial intelligence (AI) and machine learning (ML) have been applied in large datacenters and the cloud – places that have the huge amounts of storage required to hold the training datasets. This works great for applications such as business intelligence and high-speed stock trading, but for a large number of emerging edge-based applications it is extremely difficult getting this data back to the datacenters or the cloud. Just as often, the data needs to be evaluated at the edge, which at the very least means that the inference engines must be forward-deployed to the edge.

A good example of this movement towards the edge is what Verizon is doing with their AI and ML efforts. 5G certainly helps with these efforts, as does the proliferation of content delivery networks (CDNs) like that operated by Verizon. For Verizon, moving AI and ML to the edge enables them to be able to inspect cell towers with fleets of drones, which then can be applied to pipelines, bridges, and buildings! Networks to connect autonomous vehicles together as a

"single learning entity" are another example of this. These are some of the topics that Ganesh Harinath of Verizon Media will be talking about at the NVIDIA GPU Technology Conference (GTC) in mid-April.

One of the issues that Verizon Media is tackling is "how do you securely deploy AI and ML applications to the edge"? While vendors have deployed "standard" applications (and updates to those applications) to the edge through CDNs in the past, it is significantly more difficult with AI/ML-based applications, as they are "dynamic" in nature. This is why Verizon Media has developed the **Leo** AI edge platform, which is shown in the diagram. In this model, the pink boxes are deployed at the edge (think 2U servers) and perform much of the work, while the cloud datacenter components (shown in the blue boxes) provide "support" for the edge. The model is not only concerned about security and model deployment – it also focuses on keeping the entire network synchronized. After all, if multiple ML models are analyzing data in different geographical areas, the results will differ and it will be very difficult to utilize the results in an end-to-end manner. This is just one approach to deploying AI and ML to the edge, and it is something that all technology suppliers will need to figure out as the Internet of Things (IoT) generates more and more data.

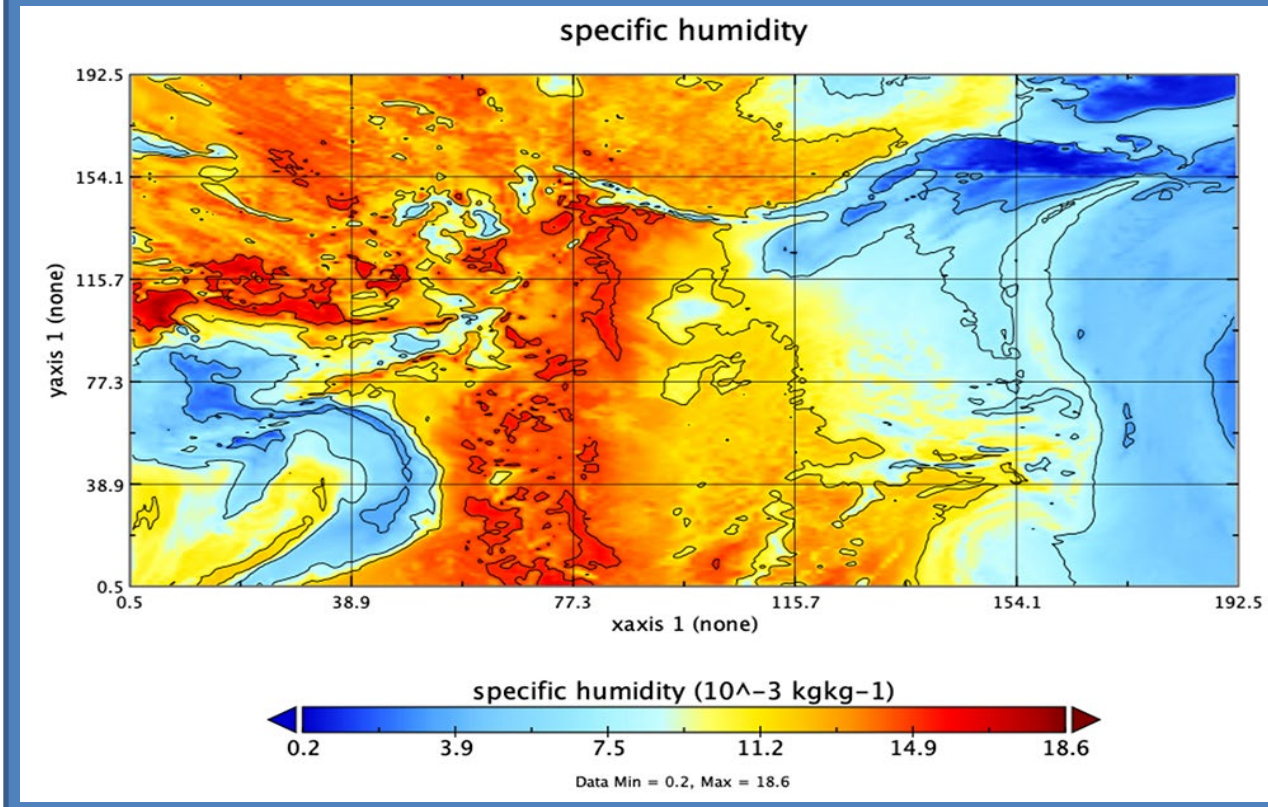


Listen to the entire conversation regarding the paradigm shifts occurring in the space as explained by Ganesh Harineth, Vice President & CTO, 5G MEC, AI Platforms & Next-Gen Applications for Verizon Media, as presented at the Red Hat OpenShift Gathering on Data Science.

Building Storage Solutions for High-Performance Computing

Kioxia, NVIDIA, Weka

High-Performance Computing (HPC) originally started out in defense and research domains, such as universities and government agencies. The workloads that the initial HPC solutions were used for included particle physics, fluid dynamics, nuclear weapons modeling, and research and aerospace projects – and, obviously, the space program. What we have seen in the last several years is HPC being applied to new problems in commercial enterprises. This includes product modeling to speed time to market, business intelligence (analyzing what customers are doing), and for artificial intelligence and machine learning. These and other areas have formed a toehold for HPC in the commercial space and enterprise, which will undoubtedly grow over time.



This visualization depicts specific humidity at atmosphere Level 54 (in one of six cubed-sphere grid tiles covering the globe) from the Joint Center for Satellite Data Assimilation (JCSDA) Finite-Volume Cubed-Sphere Dynamical Core Global Forecast System (FV3-GFS) Joint Effort for Data assimilation Integration (JEDI) 3DVar application, which was wrapped into a supercontainer for running on high-performance computing (HPC) platforms.

As you can imagine, there are a lot of differences between enterprise datacenter storage architectures and HPC storage architectures. Conventional datacenters (especially private cloud implementations) are really built to simplify the job of the IT staff by using homogeneous resources that simplify migration and scaling and flexibility. Virtualization of all resources (compute, storage, and networking) is really important to achieving application availability and data integrity, with a focus on avoiding loss of data and downtime. Performance, generally, while it can be important, is secondary to flexibility and data integrity concerns. While there are some enterprise architectures implemented for performance such as data lakes, these are outliers.

The priorities for HPC architectures are the opposite of enterprise datacenters architectures. You might have thousands or hundreds of thousands of processors and GPGPUs that are drawing data from a single storage pool. They tend to run applications as projects or batches, where a job is set up, ran, and torn down when complete. Because of that, HPC architectures are highly performant and they don't necessarily have to be standard. Being able to configure a cluster for optimum performance is usually the primary concern.

The challenges of moving HPC into the enterprise require IT staff to think very differently. You have to plan differently, manage resources differently, and, in the end, you have to execute very differently because this is a very different environment. Maximizing HPC storage performance was the subject of the G2M Research February 23rd webinar, with Kioxia, NVIDIA, and Weka.

Kioxia discussed the importance of local storage in each compute node to maximize performance by reducing storage latency, especially for extremely large datasets, and the importance of connecting storage nodes to compute nodes via NVMe-oF™. Kioxia also talked about the fit of their various SSDs for HPC architectures.

NVIDIA spoke about using InfiniBand as the interconnect within supercomputing clusters, as it is the most widely deployed interconnect in HPC due to its deterministic performance. They also discussed using NVIDIA's GPUDirect Storage (MagnumIO) interface to avoid moving data through the CPU.

Weka talked about I/O patterns and how they vary across HPC workloads, how GPUs have increased the density of HPC clusters, and why current NAS solutions cannot provide the bandwidth required to feed these clusters.

Kioxia, NVIDIA, and Weka provide approaches to storage architectures that maximize the performance of HPC Clusters. Hear the discussion from our last webinar here.

Poll Question from Our Webinar

"Can Your Servers Handle the Size of Your SSDs?"

with Industry Leaders Intel, Kioxia, and Lightbits Labs

When looking at solutions for the blast radius problem, which of these approaches has your organization explored?

(check all that apply):

Scale-Out Flash Storage (SOFS) software solutions:	37%
Distributed File Systems:	33%
Networked SSDs (Ethernet, NVMe-oF, etc.):	30%
Composable Infrastructure:	10%
Centralized storage arrays:	37%
Other:	17%

2021 Predictions from Andy Fernandez

Zerto

Andy Fernandez, Senior Product Marketing Manager for Zerto, gives his Predictions for 2021, from Storage Newsletter Vendor Predictions for 2021.

Now that we've passed the hype, the cloud has become a catalyst of digital transformation. Covid-19 stress tested our infrastructure globally, and it showed that the cloud can actually scale and support the surge in provisioning which

confirmed it as a reliable source of infrastructure. This was made possible not only because of how readily available the cloud is but also how it allows people to scale quickly, spin up new resources, and accelerate application development.

Now organizations are realizing they can move their data protection and disaster recovery services to the cloud, resulting in an increase in speed, agility, and efficiency.

In 2021, modern organizations will move even more workloads to the cloud and continue to adopt cloud-native services, specifically containers and applications for DevOps. By the end of the decade, enterprises will run most of their production environments in the cloud. Companies will move away from building new sites or buying more hardware in favor of pursuing an operational model with the cloud. In order to achieve this, organizations will need data management, protection and mobility solutions that facilitate this move, not act as an impediment.

Enterprise Storage Events – All Virtual

April 12-16 GTC21

April 18-23 Enterprise Data World

April 18-22 Cloud Computing 2021 - IARIA

April 21-22 SNIA Persistent Memory & Computational Storage Summit

April 27-28 Red Hat Summit 2021

May 4-6 Gartner Data & Analytics Summit

G2M Multivendor Webinar Series

Our February webinar was “Storage Architectures to Maximize the Performance of HPC Clusters” was sponsored by Kioxia (Matt Hallberg), NVIDIA (Reggie Reynolds), and Weka (Joel Kaufman). View the recording and/or download a PDF of the slides.

Our 2021 webinar schedule! Click on any of the topics to get more information about that specific webinar. Interested in Sponsoring a webinar? Contact G2M for a prospectus.

G2M RESEARCH

March 23:	One Year after COVID-19: How Did Storage Architectures Perform for Biotech AI Modeling & What Can We Learn From This?
April 20:	The Race to be Relevant in Autonomous Vehicle Data Storage (both On-Vehicle and Off-Vehicle)
May 18:	Responsive and Efficient Storage Architectures for Social Media
June 15:	It's 2021 - Where Has NVMe-oF™ Progressed To?
July 13:	Computational Storage vs Virtualized Computation/Storage in the Datacenter: "And The Winner Is"?
Aug 17:	AI/ML Storage - Distributed vs Centralized Architectures
Sept 14:	Composable Infrastructure vs Hyper-Converged Infrastructure for Business Intelligence
Oct 12:	Cloud Service Providers: Is Public Cloud, Private Datacenter, or a Hybrid Model Right for You?
Nov 9:	The Radiometry Data Explosion: Can Storage Keep Pace?
Dec 14:	2021 Enterprise Storage Wrap-up Panel Discussion