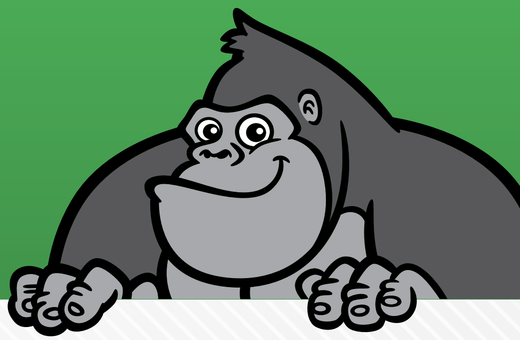
ActualTech Media

SCALITY

Hewlett Packard Enterprise

THE GORILLA GUIDE TO...®

EXPRESS EDITION



Object Storage for Data at the Edge

Joey D'Antoni

Inside the Guide

- The Three C's of Edge Computing
- Moving Data from the Edge to the Cloud
- Object-Based Storage for Edge Computing

Object Storage for Data at the Edge

Express Edition

By Joey D'Antoni

ACTUALTECH MEDIA

6650 Rivers Ave Ste 105 #22489
North Charleston, SC 29406-4829
www.actualtechmedia.com

PUBLISHER'S ACKNOWLEDGEMENTS

EDITORIAL DIRECTOR

Keith Ward

DIRECTOR OF CONTENT DELIVERY

Wendy Hernandez

CREATIVE DIRECTOR

Olivia Thomson

SENIOR DIRECTOR OF CONTENT

Katie Mohr

PARTNER AND VP OF CONTENT

James Green

ABOUT THE AUTHOR

Joey D'Antoni is a Senior Architect and Data Platform MVP with over a decade of experience working in both Fortune 500 and smaller firms. He is currently Principal Consultant for Denny Cherry and Associates. He is frequent speaker at major tech events, and blogger about most technology topics. He believes that no single platform is the answer to all technology problems. He holds a BS in Computer Information Systems from Louisiana Tech University and an MBA from North Carolina State University.

ENTERING THE JUNGLE

Introduction: Edge Computing 7

Chapter 1: Edge Computing Scenarios 9

- The Three C’s of Edge Computing 10
- Categories of Edge Computing 10
- Types of Edge Computing Devices 12

Chapter 2: Edge Data Scenarios 13

- Storage Management at the Edge 15
- Storage and Data Durability 16
- The Impact of Storage Latency 19
- Measuring Storage Performance 19
- Moving Data from the Edge to the Cloud 19

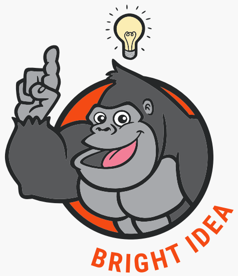
Chapter 3: Object Storage for Edge Computing 21

- Delivering High Performance 22
- Data Availability and Durability 23
- Hybrid Cloud Scenarios 23
- Storage Management 24
- The HPE-Scality Advantage 24

CALLOUTS USED IN THIS BOOK

The Gorilla is the professorial sort that enjoys helping people learn. In the School House callout, you'll gain insight into topics that may be outside the main subject but are still important.

This is a special place where you can learn a bit more about ancillary topics presented in the book.



When we have a great thought, we express them through a series of grunts in the Bright Idea section.

Takes you into the deep, dark depths of a particular topic.

Discusses items of strategic interest to business leaders.

ICONS USED IN THIS BOOK

DEFINITION
Defines a word, phrase, or concept.

KNOWLEDGE CHECK
Tests your knowledge of what you've read.

PAY ATTENTION
We want to make sure you see this!

GPS
We'll help you navigate your knowledge to the right place.

WATCH OUT!
Make sure you read this so you don't make a critical error!

TIP
A helpful piece of advice based on what you've read.

Edge Computing

Welcome to The Gorilla Guide To...® Object Storage at the Edge, Express Edition! Edge computing has become a prevalent part of systems architectures across a wide array of industries. A conflation of smaller, more powerful computing devices, the Internet of Things (IoT), and the demand for data-driven processes have increasingly led to the digitalization of real-world processes. A couple of simple examples of how ordinary processes have become digitized include my two favorite hobbies—cycling and cooking.

When I go for a bike ride, I have a computer that records my speed, distance, power output, cadence, and many other environmental data points. Beyond this, my bicycle also has computer-integrated electronic shifting—meaning each gear shift is recorded and integrated with the data from the ride.

When I barbeque, I have a thermocouple device that records the temperature inside my smoker and adjusts the airflow to the burning wood to manage the cooking process. While these are just two examples of digitalization processes, the point is that any sort of automation demands both data collection and processing.

Traditional computing platforms weren't designed with edge computing in mind. Servers are designed to run in data centers and assume centralized control platforms. Historically, disk storage has been even more fragile and required even tighter environmental controls. Hard drive components like platters and heads are extremely sensitive to dust and vibrations.

One of the substantial changes in the storage industry has been to move from mechanical spinning disks to solid-state drives (SSDs) that have no moving parts. SSDs offer large performance gains and the benefit of being much lighter. An SSD can weigh between 10% to 50% less than a similar hard disk drive. Beyond just weight, SSDs have lower power and cooling requirements, which means they can be packaged into smaller devices with longer battery life.

Edge computing device management can be challenging, much like any computing device that doesn't run in a fully controlled environment or isn't easy to access physically by the IT personnel who may need to respond to errors or hardware problems.

Keep in mind that when working with field devices, transient and minor hardware failures will occur. You need the ability to integrate with monitoring solutions so that if there are hard failures, replacement devices or hardware can be dispatched to the field.

This Guide will take you through the challenges of edge computing, and how to meet those challenges. One of the biggest challenges is storage, and specifically how object storage fits into this new world of highly distributed applications, data, and infrastructure.

Object storage is increasingly crucial for edge environments, and this book will also reveal how advances in object storage technologies are enabling faster, more efficient edge computing, giving companies a significant competitive advantage.

So if you're ready, let's get started on the journey. We'll begin with a look at how edge computing breaks down into different classifications.

CHAPTER 1

Edge Computing Scenarios

Edge computing has a nearly bottomless pit of use cases. These scenarios have differing compute and storage requirements, depending on the data type collected and where these devices sit in the business data flow.

In addition, there's a clear demand for setting up real-time analytics at the edge. This creates a problem, as there are a couple of limitations that impact edge architecture, including limited bandwidth from the edge to the hub.

Because of this network latency challenge in getting data back to a central hub, the ability to provide real-time analytics can improve operational efficiency, deliver better security, and enable several improved business outcomes.

> **The definition of edge computing can be fuzzy.** For this Guide, we'll define it as locating compute resources closer to the devices using them. This places the edge firmly outside the traditional data center into all kinds of untraditional places where access and bandwidth can be sketchy.

The Three C's of Edge Computing

The easiest way to think about "the edge" is as a place. This could be a manufacturing floor, your house, an office campus, a wind farm, a power plant, or even a ship at sea. Beyond simple telemetry streaming applications, the "intelligent edge" has systems in place to perform computation and control processes at the edge device. Three components make up this intelligent edge:

- Connectivity: Permanent or intermittent connection to the computing device at the edge. This connectivity can be Wi-Fi, wired, or cellular. Larger bandwidth allows for richer data interchange between the edge and the data center or cloud.
- Compute: Compute capacity provides the ability to perform calculations and add intelligence to edge devices. The trade-off to adding compute power to devices is that their power demands will typically increase.
- Control: Controlling processes are where edge devices come together to provide value to processes. An example of this could be a car's engine that gets reduced power to protect the engine after a sensor fault is detected.

While the three C's are the core elements of the intelligent edge, these technologies are implemented in different ways for different industries and use cases.

Categories of Edge Computing

Edge computing is commonly broken down into a few different categories of use-case types. These use cases often map to different kinds of hardware and computing resources. The categories shown here aren't industry standards, but loose classifications instead:

- Cloud
- Compute edge

- Device edge
- IoT/sensor edge

You should understand the different use cases for each.

Cloud edge: The cloud refers to hyperscale data centers operated by providers such as Amazon, Google, and Microsoft. The cloud can be a host for edge computing, or it can be a hub for other distributed edge devices.

Compute edge: The compute edge consists of more traditional hardware and storage containing one or more racks of servers and disks. These are located close to IoT devices and provide storage and processing capacity to those devices. The benefit of having compute hardware close to the edge is that it typically provides higher levels of compute and network bandwidth as compared to device-only edge configurations.

Device edge: The device edge consists of just one to a few servers, and can also be described as a "nano data center." Servers and storage in this category are typically designed to run without traditional data center cooling and are typically found in locations like factories, power plants, or smart cars. These devices can provide more computing power than sensors, but are limited in their power and capacity.

IoT/sensor edge: IoT includes devices like smart lightbulbs, security cameras, or thermostats. These devices have minimal compute or storage capacity, and typically have enough storage only to cache data until it's stored in an upstream permanent location.

Understanding these case functions can help you determine your computing requirements, enabling you to better choose your infrastructure for various edge computing scenarios.

Types of Edge Computing Devices

As you've seen, there are a wide variety of edge computing architectures and patterns. To support these patterns, diverse computing devices may be employed. These edge computing devices go well beyond normal PC hardware, and are driven by increases in edge computing that are causing exponential data growth.

A study from Seagate and IDC suggests that by 2025 the amount of data generated annually will increase to 175ZB, largely driven by increases in data at the edge (see **Figure 1**). However, network bandwidth to edge devices is not increasing at that same rate, which means more and more of that data will need to be processed and analyzed at the edge. This will ultimately drive up storage and compute requirements.

The type of workload each edge scenario supports will dictate the amount of storage needed both locally and at the hub. There are two main storage patterns at the edge: edge storage and edge cache.

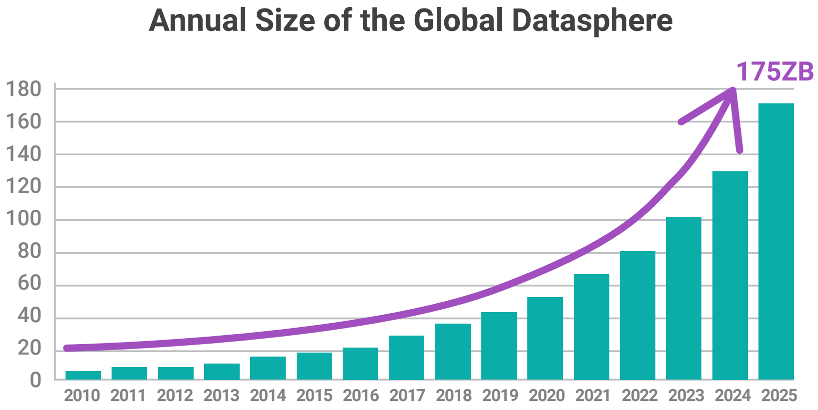


Figure 1: The rapidly growing size of data (Source: Seagate/IDC)

CHAPTER 2

Edge Data Scenarios

As mentioned, the ever-increasing volume of data at the edge typically requires some degree of local processing, followed by persistent data storage for longer-term analysis.

A very simple example of this would be checking the temperature of an engine. Unless the temperature is above a warning threshold, it can be ignored—at least in the short term. However, you might want to perform a deeper analysis of the variance in engine temperature in relation to other environmental factors over time, and for that, you need the complete set of values. This common pattern is what's known as Lambda architecture, and it can take a variety of different shapes.

Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch and stream-processing methods. This approach to architecture attempts to balance latency, throughput, and fault-tolerance by using batch processing to provide comprehensive and accurate views of batch data, while simultaneously using real-time stream processing to provide views of online data. The two view outputs may be joined before presentation. The rise of lambda architecture is correlated with the growth of big data, real-time analytics, and the drive to mitigate the latencies of map-reduce. (Source: Wikipedia)

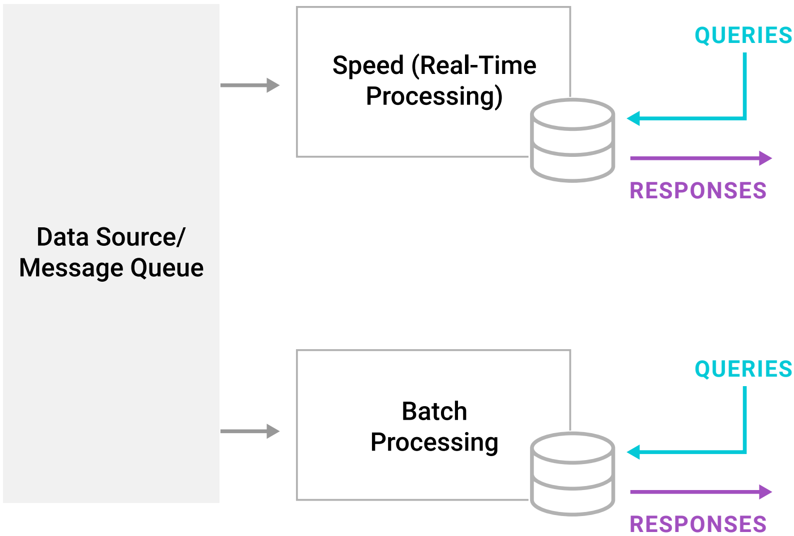


Figure 2: Lambda architecture design pattern

Lambda architecture (see **Figure 2**) provides a methodology for breaking data processing into two separate streams. Typically, the speed layer will be some kind of a streaming query engine—the open source Kafka being a prominent example—triggering an action for any out-of-band values.

In the simple engine temperature example, if the values of three consecutive records are above the safety threshold, an alert would flash to the operator. These alerts are typically logged, but the actual records do not persist in the speed layer.

The batch processing layer, on the other hand, allows for longer-term analysis and may pay attention to minor variations in data trends over time. Since you're capturing all records—not just the outlier—your data storage needs increase dramatically.

As data volumes increase, you will typically move more processing power to the edge, usually resulting in summarized and compressed data being sent back to reduce bandwidth requirements. This processing increases the demand for storage performance and reliability at the edge location.

Storage at the edge needs to have several characteristics, including the ability to potentially run on battery power, the ability to be more power-efficient, and the ability to tolerate harsher temperatures and vibration conditions.

Beyond just physical characteristics, storage controller software also needs to include redundancy to tolerate single drive failures and provide data durability. This means storage subsystems must have both software intelligence and hardware redundancy to offer high levels of data resilience, while providing rapid recovery from any failures to provide consistent access to data.

Storage Management at the Edge

Since edge locations are inherently decentralized, they must provide a common framework for management and analytics. While the storage and compute devices are decentralized, the people managing these systems are typically centralized and prefer to manage their storage in a consolidated fashion. Managing tens or hundreds of different devices individually simply isn't practical.

Ideally, these solutions can integrate with common public or private cloud storage infrastructure like Amazon Simple Storage Service (S3) or Azure blob storage. These are the typical long-term storage endpoints for most edge data, even in private cloud scenarios.

Another part of monitoring both storage and storage infrastructure that modern systems should adopt is the ability to be software-defined. Software-defined storage has become a given in the cloud world, where everything can be defined in a YAML or JSON file and deployed to an API via a DevOps pipeline. This software enablement allows for both common management tooling and for the deployment of more automation.

JavaScript Object Notation (JSON) and Yet Another Markup Language (YAML) are two ways of storing object data. Both commonly used with software-defined infrastructure solutions. They're similar in that YAML is a superset of JSON, but JSON is more explicit and strict in its format than YAML. YAML is designed to be human-readable. Typically, applications will write data into JSON, whereas humans will write YAML code.

Another aspect of storage management that must be considered is tiering. While there has always been the notion of tiered storage—faster and more expensive storage for frequently accessed data, denser and slower storage for archival data—tiering is something that has accelerated in cloud storage solutions for cost management. Your storage tooling should support tiering and transfers for storage lifecycle management.

Storage and Data Durability

Older storage systems relied on RAID, or Redundant Array of Independent Disks, to provide durability. RAID allowed for a group of disks to tolerate the failure of one or more drives, but was prone to failure if there were simultaneous disk failures or errors during re-build operations, which substantially risked the loss of data. RAID has largely been supplanted by erasure coding in modern environments, and nearly all current object-based storage technology depends on erasure coding to provide data durability.

Erasure coding technology

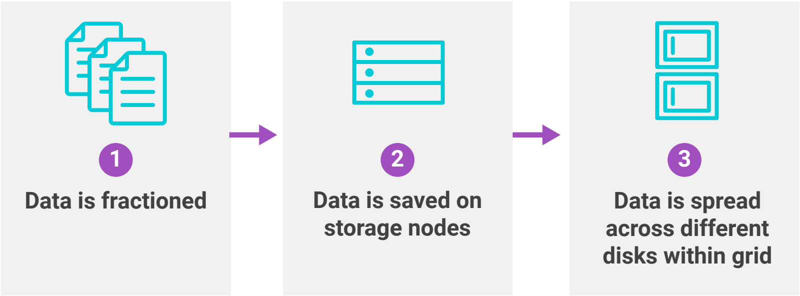


Figure 3: Erasure coding visualized

Erasure coding is a method of data protection where data is broken into fragments and then expanded. The data is encoded with redundant data pieces and stored across devices in an array. In the event of a failure, the data can be reconstructed from data segments stored on other drives in the array. Erasure coding is more granular and efficient than RAID.

Storage Performance

Frequently, administrators only think of storage in terms of overall capacity, and not in terms of storage performance. Storage performance is measured in two dimensions that are correlated:

- IOPs, the number of read and write operations that can be performed per second
- Latency, how long a given storage operation takes to complete

The other important factor for most storage is the amount of network bandwidth from the computer or edge device to the storage, which is important for object stores that are typically network-connected. If there isn't enough bandwidth, that network can act as a bottleneck for the entire application.

When trying to identify a single metric to gauge storage performance, your best bet is to look at the latency. More specifically, it's important to look at your *maximum* latency and not just the *average* latency.

For workloads that have very high peaks, their maximum latency can significantly exceed the average, as the large amount of lower latency values can make the average look healthy, even though there's significant latency at the maximum. These spikes can indicate a lack of adequate performance to meet your application requirements at peak workload.

When you're scaling your storage at the edge, it's important to consider not only the raw volume, but also the performance capacity of the storage. The type of storage becomes a crucial factor here, and object storage often emerges as the best solution.

One benefit of object storage is its ease of scaling horizontally by adding additional nodes that offer improved performance and storage capacity at a linear rate. The ability to adapt to new nodes, as well as the ability to balance storage operations across the newly added nodes, is a key element for any storage solution (more on this in the next chapter).

Scaling your storage capacity and performance should be easy to do technically, as well as predictable in terms of cost. To meet this goal, your storage solution needs to be able scale easily, and affordably. Linear scale is what you're looking for, particularly as your data volumes grow beyond petabytes.

The Impact of Storage Latency

Storage latency can lead to many problems within your application stack. When your application is waiting on data, it's simply idle until it receives the data requested or gets an acknowledgment that a write operation was completed.

If your storage is slow, your edge application may not be able to make necessary decisions, resulting in the failure of your application goals. Modern storage platforms like NVMe allow storage to operate at extremely low latencies—well below the former gold standard of a single millisecond. You may see peaks of 10 to 20 milliseconds, but anything beyond indicates a lack of scale or misconfiguration. Optimizations from your storage provider and leveraging flash caches for object store indexes to provide rapid access to reads and writes are ways to overcome the scaling issue.

Measuring Storage Performance

Beyond just storage performance and capacity, your storage appliance should provide insights into performance to allow administrators to understand their performance and storage utilization.

To that end, monitoring should go beyond simple up/down and IOPs—it should highlight capacity and object counts to show utilization. Ideally, monitoring and analytics data should be accessible via APIs so that it can easily integrate with other centralized tooling. The tooling should also provide both capacity planning and alerting mechanisms that keep you in control of of your capacity.

Moving Data from the Edge to the Cloud

One of the inherent limitations of edge computing is physical space—which in turn means limits on the amount of storage capacity available to edge devices. Since the data collected at the edge can provide business value through its overall analysis, it's important to have a methodology to store that data in a central store.

One common approach many organizations are taking toward that goal is to move cold edge data into a cost-effective cloud-based object store.

This cloud store can be in a public cloud like AWS or Azure, or your own private cloud. One of the benefits public cloud providers have built into their storage platforms is lifecycle management, which also auto-tiers data between hot, cold, and archive tiers in the cloud.

This can reduce overall storage costs, particularly for data that's untouched for long periods of time, or data being held for regulatory reasons. Some private cloud solutions provide similar functionality.

To reduce management overhead, you should ensure that your storage solution supports automatic tiering of older data into this cloud tier. Especially as your number of edge devices increases, it's simply not practical to manually administer this functionality. In addition, rudimentary automation techniques, like shell scripts to move the data, fall well short of what's necessary. Your storage solution should have an auto-tiering option, which should ideally work with public cloud lifecycle management.

CHAPTER 3

Object Storage for Edge Computing

Object storage helps solve the problems faced by ever-increasing storage volumes. Instead of splitting disks into individual blocks, object storage stores blobs, which can be files, or parts of files, with an expandable amount of metadata. This metadata is associated with the object and a globally unique identifier, which allows the object to be found across a distributed system.

This identifier allows the data to be found in different parts of a data center, or even across different data centers. Object storage use has been driven by cloud platforms like Amazon Web Services with S3, and Microsoft's Azure blob storage.

Beyond the clouds, big data technologies like Hadoop are dependent on object-based storage. It has become the primary target or many applications, including:

- Enterprise backup and archive
- Big data analytics
- Media content distribution

These applications are both performance- and capacity-intensive, and object stores have evolved to meet both challenges.

Beyond providing nearly unlimited scale, object-based storage provides a number of advantages for edge scenarios. For instance, object storage allows for gradual and seamless growth and simplified management of the storage solution.

In addition, data durability is enhanced through the use of erasure coding, while delivering the highest performance possible with the lowest overhead.

Finally, object storage can save significant money with policy-based tiering. That means moving data around according to its use or value, and since the majority of most organizations' data isn't "hot," it can be directed to less expensive object storage, ready when you need it, but not costing you tons of money by sitting in a premium storage tier.

Delivering High Performance

Object storage takes advantage of newer storage technologies such as solid-state disks and NVMe to deliver best-in-class performance.

Improving performance isn't as simple as just adding faster storage—you need to ensure that the network to your storage has adequate bandwidth to meet the data flows delivered by the storage.

Beyond the raw storage, the storage management infrastructure needs to supply high-performance API calls and leverage caching, which can reduce the overhead for normal read and write operations.

API accessibility is another key point of most object-based storage models. Since the store isn't accessed by the operating system, interaction with the storage occurs through Rest APIs at the application level. While there are several different implementations of these APIs, the most popular one is Amazon S3. Because of its broad adoption, many vendors have emulated the Amazon S3 API for storage accessibility. This allows programs written for Amazon S3 to be used against a wide variety of storage solutions in public cloud and private use cases.

Data Availability and Durability

Like RAID, object-based storage may maintain data availability by keeping copies of the data. In some cases, this data can be replicated to another region asynchronously, providing an additional data protection measure.

This storage redundancy is coupled with erasure coding to provide additional levels of data protection, ensuring that modern object stores have extremely high levels of durability.

One of the important measures of your object storage, especially in edge computing paradigms, is the amount of overhead incurred by your erasure coding solution. The more efficient the erasure coding is, the more available storage you'll have from your object store.

Hybrid Cloud Scenarios

The growing combination of data volumes, along with the limited capacity of edge storage, means nearly all edge solutions are hybrid solutions. This means the storage at the edge, while supporting the processing and alerting of operations, serves as a holding area before the data is pushed into a centralized object store, which can then function as either a data lake or simply storage. There are a few different scenarios in terms of how these application models are built.

In most of these scenarios, the cloud—whether it be public or private—provides for large storage volumes and compute capacity to meet the larger processing requirements.

An example of this would be a smartwatch. The watch itself has limited processing capacity, while the watch's full functionality is driven by a mobile application that has both computing capacity on the phone and in the cloud.

The major benefit of these hybrid models is flexibility—you can move data to different locations based on time requirements, performance needs, or security and regulatory essentials.

Storage Management

Beyond the performance and capacity of your storage, management tools are an important part of your storage infrastructure. You want to ensure your solution provides a single place to manage your storage—whether it be local or at an edge location. Beyond just providing a single pane of glass for management, solutions that are API-accessible can integrate with a wide array of monitoring and alerting systems, which can better integrate into your automation workflows. Beyond management activities, having consumption and performance information easily available is a major benefit to DevOps teams.

The HPE-Scality Advantage

By now, it should be clear that object storage and edge computing go together like peanut butter and jelly. They complement each other, and the incredible promise of the computing edge can be realized with object storage.

Realizing that potential, companies like Scality and HPE have gone all in on object storage for the edge. For instance, the Scality ARTESCA solution provides robust performance and full support for all the major storage APIs to support your edge computing needs.

ARTESCA is an object-only storage solution with ultra-data durability that can support exabytes of data. ARTESCA also provides comprehensive storage analytics in a metrics and reporting dashboard that's also fully accessible through a Rest API.

ARTESCA provides seamless scale-out by simply adding new nodes into the cluster and adding both storage capacity and performance. It comes in a number of deployment options, from one server to a nearly unlimited scale-out model with 60TB per node.

HPE and Scality have co-developed a full portfolio of ARTESCA platforms, enabling both fast and single-node edge object systems, making it easy to deploy object storage everywhere.

In this Gorilla Guide you've learned about different ways companies are using edge computing, how object-based storage can help you meet your edge computing needs, and what elements you need to look for in an edge storage solution. Now it's time to get started with Scality ARTESCA.

ABOUT SCALITY & HPE

Scality® storage propels companies to unify data management no matter where data lives—from edge to core to cloud. Our market-leading file and object storage software protects data on-premises and in hybrid and multi-cloud environments. With RING and ARTESCA, Scality's approach to managing data across the enterprise accelerates business insight for sound decision-making and maximum return on investment. To compete in a data-driven economy, IT leaders and application developers trust Scality to build sustainable, adaptable solutions. Scality is recognized as a leader by Gartner and IDC. Follow us @scality and LinkedIn. Visit www.scality.com, or subscribe to our company blog.

Hewlett Packard Enterprise is the global edge-to-cloud platform-as-a-service company that helps organizations accelerate outcomes by unlocking value from all of their data, everywhere. Built on decades of reimagining the future and innovating to advance the way people live and work, HPE delivers unique, open and intelligent technology solutions, with a consistent experience across all clouds and edges, to help customers develop new business models, engage in new ways, and increase operational performance. For more information, visit: www.hpe.com.

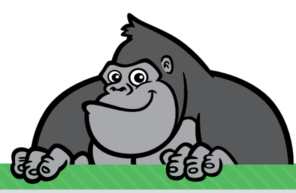
ABOUT ACTUALTECH MEDIA

ActualTech Media

ActualTech Media is a B2B tech marketing company that connects enterprise IT vendors with IT buyers through innovative lead generation programs and compelling custom content services.

ActualTech Media's team speaks to the enterprise IT audience because we've been the enterprise IT audience.

Our leadership team is stacked with former CIOs, IT managers, architects, subject matter experts and marketing professionals that help our clients spend less time explaining what their technology does and more time creating strategies that drive results.



If you're an IT marketer and you'd like your own custom Gorilla Guide® title for your company, please visit https://www.gorilla.guide/custom-solutions/