

# Machine Data

## Manage the Crushing Volume, Variety and Velocity of Multi-Petabyte Scale Machine Data

Machine data – from IT systems, deployed devices and embedded sensors – holds the promise of ever-greater insight, but only for those who successfully ingest, process and harness the flood of information. From life and earth sciences to media & entertainment to automotive, companies everywhere are turning to Qumulo for the scalability and real-time analytics needed to embrace the promise of multi-petabyte scale machine data.

### The Three Ages of Data Storage

Traditional IT infrastructure was designed around data which was predominantly created by humans, from email and documents to business transactions, databases and records. Twenty to thirty years ago, the volume of data was primarily driven by business processes in the form of online transactions. It was for things like a bank generating customer statements or similar documents. Those transactions were conducted through mainframes, stored in traditional databases, transferred across storage area networks and related infrastructure until they ended up in your mailbox. Many familiar names in IT made their name during this First Age of Data.

Then a Second Age of Data arose, still human-centric but driven less by processes and more by an explosion in content: office documents, streaming audio/video, digital imaging and photography, email, websites, etc. The addition of this variety of file types, formats and sizes on top of traditional data volume soon led to a huge increase in storage requirements. Pioneering vendors from the previous transactional age were subsequently replaced by scale-out companies that could meet the need of scale.

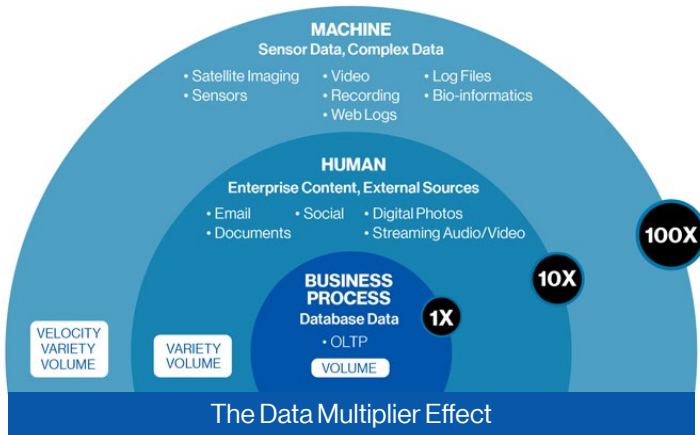
### Volume, Variety and Velocity of Machine Data

Now, in the Third Age of Data, industries are faced with a rising tide of data being generated from machines – sensor data, imaging, data capture, logging or monitoring, and more. This hyperscale growth in machine-generated data provides a wealth of opportunities for enterprises to find new insights from complex systems. It is valuable only for those that can efficiently understand and manage multi-petabyte scale data storage and processing.

At its core, the challenge of machine data involves dealing with new parameters: a previously unheard-of volume and variety, now compounded by the rapidity and frequency with which machines generate multi-petabyte scale unstructured data. This volume, variety and velocity create a “data multiplier effect” that can translate into orders of magnitude inflation in the scale of the data being collected.

### Key Benefits for Storage of Machine Data:

- Improves IOPS performance to handle growing volume, variety and velocity of data
- Provides real-time data analytics for instant file-level visibility and control regarding usage, capacity and performance
- Saves valuable IT time through reduced management and reporting
- Increases transparency and provides opportunity to better harness data thanks to intuitive dashboard and REST-based API
- Delivers proactive support for uninterrupted operations and fast root-cause resolution



### Machine Data Volume

Embedded sensors in automobiles and roadways supply information about location, speed, direction and operation, allowing for everything from better traffic management to vehicle monitoring, routing and entertainment. Similarly, network packet, traffic, call and log monitoring provides insight into service operations and security, keeping IT datacenters or telecommunications networks safe and sound. The scale of this sensor and packet data is massive and growing every year.

### Machine Data Variety

Then there's the variety of the data at hand. In some industries, companies rely on constant tiny measurements from ground or equipment-mounted sensors and devices, as well as huge and complex satellite imagery, weather models, geospatial data, etc. Similarly, many companies have a mix of data sources (different machine systems, human, etc.), with corresponding differences in data size and type. In either case, systems and storage optimized for one end of the spectrum may not be able to readily handle the other.

### Machine Data Velocity

Finally, there is the velocity of machine data. Sensors, satellites, networked systems and connected vehicles all have one thing in common: they never sleep. These machines typically operate on the basis of continual measurement (24x7x365), constantly streaming data for processing and storage. Moreover, the flood of data can quickly spike. In life sciences, as an example, large-scale systems or teams rapidly generate tens of millions of files – or multi terabyte-size models – in just a few hours. Keeping up with that data load, and more importantly, understanding its constant ebb and flow, is an equally massive challenge.

## Multi-Petabyte Scale Machine Data Processing and Storage

The challenge of machine data is the challenge of ingesting, processing and harnessing the information, at scale, to gain real-world insight: How do you manage the onslaught of data? What data should you keep? What data should you delete? How do you process the raw or intermediate data so that you can understand it and gain actionable insights? And, how do you translate those insights into tangible business value?

While machine data may grow to be a flood, it often starts as a seemingly manageable stream. In the past, IT teams have addressed that trickle by using network attached storage (NAS) to create a central storage repository—ideally one with a scale-out design that could provide for expansion without the need for multiple namespaces.

But, as that stream of machine data becomes a torrent, analyzing and managing the data becomes increasingly difficult. The problem becomes distinguishing hot from cold data, understanding capacity and performance impact by user or application, and precisely forecasting when capacity will run out. Running reports and analysis using legacy scale-out NAS technologies can take days, weeks or sometimes months, creating both an endless cycle of reporting which can impact system performance and assessments that are out-of-date long before they're even complete.

It quickly becomes impossible to accurately assess larger network or data processing issues. Is a given HPC node slow because of the storage? Or is the link saturated? Who knows?

Different workflows create different stresses on the infrastructure, and the best storage solution must elegantly handle those varying requirements. Below are two real world workflow examples from the automotive and life sciences industries. In Figure 1: Life Sciences the workflow primarily stresses the storage infrastructure during processing and output, as custom tools for disease modeling and forecasting create a massive ebb and flow in machine data. Contrast that with Figure 2: Automotive, where the volume and velocity of source data impacts storage on ingest as well as output and analysis.

FIGURE 1: Life Sciences Machine Data Workflow

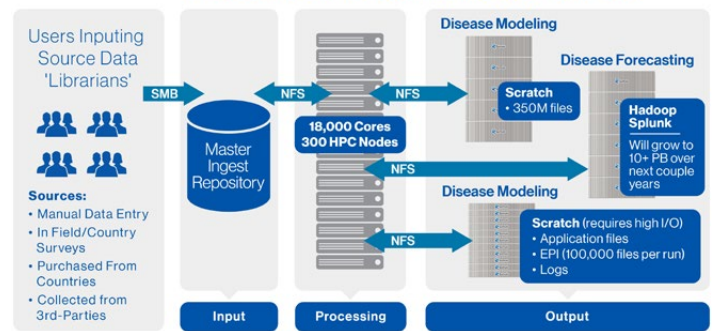
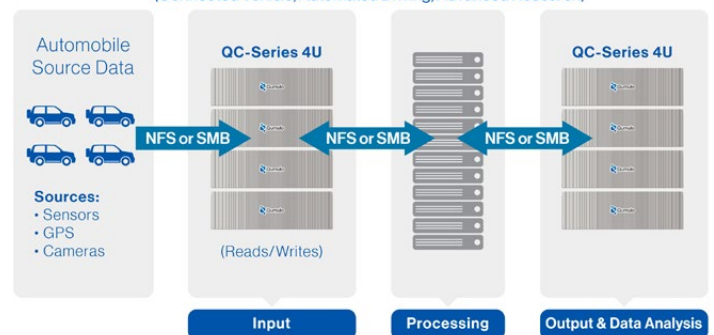


FIGURE 2: Automotive Machine Data Workflow (Connected Vehicle, Automated Driving, Advanced Research)



If traditional or legacy scale-out NAS solutions are incapable of handling the crushing volume, variety and velocity of this machine data, what new approach to data ingest and processing will deliver on the promise?

## A Modern Approach to Scale-Out Storage

Qumulo is a leader in delivering scale-out storage. Qumulo offers a modern approach to scale-out storage, delivering fast, flexible and highly scalable storage together with the real-time analytics necessary for visibility into data usage and performance at petabyte scale.

This powerful combination has a profound impact in managing the onslaught of machine data. On the one hand, it establishes new levels of performance, scalability, efficiency and reliability. On the other, for companies wrestling with scale, Qumulo Core provides the real-time visibility and control that's crucial to make accurate decisions and assessments in moments. That insight comes not only in the form of intuitive management dashboards, but also through a REST-based API that can feed information into other systems and software.

## Qumulo and Machine Data Volume

Qumulo Core is a software-only, flash-first hybrid design that leverages the price/performance of commodity server hardware, coupled with the modern technologies of flash, virtualization and cloud to deliver the performance, efficiency and scalability needed to handle the volume of machine data.

Qumulo meets volume requirements by using the highest capacity drives on the market today (10TB HGST Ultrastar He10 drives) and scaling to handle multi-petabyte deployments. Rebuild times of less than one hour deliver uninterrupted operations at scale. Qumulo provides the equivalent of +4 system protection and scale, but using only a +2 design, reducing protection overhead and improving performance on large scale unstructured data deployments.

Qumulo's real-time analytics deliver visibility and control regardless of scale. With Qumulo, companies can instantly see performance workload, current activity, "hot" files or volumes, capacity usage and accurately forecast capacity trends.

## Qumulo and Machine Data Variety

The variety of machine data – from tiny sensor measurements to giant images, forecasts or models – requires a single system that is optimized for efficiently handling tens of billions of small and large files. Qumulo Core is also flexible enough to tackle both the sequential and random access patterns involved with these different file sizes and workflows.

## Qumulo and Machine Data Velocity

The velocity of machine data impacts not only the performance of the storage itself, but also the speed with which IT teams can react to changes in the data. The high performance characteristics of Qumulo's solution are critical in handling the velocity of data ingest and processing without creating a bottleneck. Equally important, however, is the data visibility and control that allows IT teams to quickly drill down, identify and address potential issues in capacity, performance and more. It is very difficult to efficiently manage crucial data without real time visibility.

## Insight at Multi-Petabyte Scale

The age of machine data holds tremendous promise and opportunity across a broad range of industries, but only for those that have the ability to gain insight at multi-petabyte scale. Find out more about how Qumulo's modern scale-out storage can help you manage this onslaught at [qumulo.com](http://qumulo.com).

**Contact Qumulo** to learn more  
about our machine data solutions:  
**1-855-478-6856 | [info@qumulo.com](mailto:info@qumulo.com)**

