MITSUMI Distribution

HITACHI
Inspire the Next

Unstructured Data Is a Treasure Trove of Value: Unlock It With an Object Store

By Paul Lewis, Global CTO, Hitachi Vantara

POINT OF VIEW

Drive Value With Intelligent Data

Tremendous value is contained in the unstructured data hidden in dark data siloes littered through your organization. This data is poorly governed, poorly catalogued and the result of thousands of hours of work on the part of your people. The monetization of data is driving new value streams from our traditional structured data sources, but what insights are held in unstructured data sources that are harder, or currently impossible, to query?

With 80% of corporate data being unstructured and growing by over 50% a year, it is no wonder why this topic has reached the top of the priority list for CIOs. IDC data shows that the average knowledge worker spends 2.5 hours a day, roughly 30% of their time, searching for information. Honestly, I am not sure how people find things so quickly.

While productivity is certainly a benefit, many others feel unstructured data potentially represents a dangerous hidden threat to their digital livelihood based on security breaches, data leaks and attacks on privacy and compliance measures. There is no arguing that better control and visibility over data can help mitigate these risks.

One of the greatest opportunities may lie in what you don't know. What insights could be uncovered by combining and analyzing data? What decisions would have been made differently if you had more accurate data? What new revenue opportunities or intellectual property could be identified? To find out these answers and more, you need a tool that removes the barriers between data silos, wraps your data with intelligence and allows you to take your data value and accessibility to the next level. Object storage is that tool.

Where It All Started

To better understand the potential value of object storage, let us start with a few definitions. Unstructured data is really any data that does not conform to a defined or organized data model or structure, which makes them difficult to search and analyze. Some popular examples of unstructured data include emails, text files, photos, videos, audio files, webpages, presentations, multimedia, call center transcripts/recordings, financial statements, claims documents, CAD/CAM, medical imaging, and the list goes on. In contrast, databases are a very good example of structured data. Data is organized and can be programmatically searched and queried with ease. Object storage adds programmability to your unstructured data storage through metadata. Metadata is data that is added to the original file and saved together to form an object. It allows us to find the data, classify it, govern it and programmatically address it. With unstructured data, you can also do a full text index and enrich the data to make it self-identifying. Self-identifying data is data that retains its value outside of the application that created the data.

So you must be wondering why all this matters? Object storage got its foothold by addressing archiving, compliance and governance use cases with its innate ability to execute data management activities that were predicated on metadata. As a simple example, think of a legal team that needed to find and put a hold on all the files created by employee #1234 on a particular date, that mentioned a certain keyword. The metadata coupled with data management policies created a winning combination for tackling compliance and governance regulations. For most archiving use cases, object storage was typically a second tier that did not require superior performance. This was the sweet spot for object storage over a decade ago.

Object Storage Has Matured to Enterprise-Ready

80%

of organizations surveyed worldwide believe object storage can **support their top IT initiatives**

OBS is now the **top choice** for **long-term retention of unstructured data**

Object storage, originally used for Tier-2 or archival storage, is now moving to support Tier-1 workloads like security and enterprise resource management.

38% of organizations believe OBS offers the scalability needed

35% of organizations believe OBS offers the ability to analyze unstructured data and improve data quality

Source: IDC Object Storage: Foundational Technology for Top IT Initiatives

Where It's All Going

The robust data management and metadata capabilities of object storage were light years ahead of their times. Simple metadata turned to robust metadata through the ability to associate multiple annotations, or separate pieces of metadata, with a single file. Healthcare is the easiest example to showcase this, so think of an X-ray. The X-ray has system metadata from the application, the date, patient name, etc., it could also contain information about the physician, diagnosis, treatment and outcome as separate files. In this example, it's easy to see how metadata can quickly become more valuable than the file itself in terms of analyzing the metadata for the most successful treatments and such. But think about what else you can do with metadata.

As organizations look to become data driven, they need centralized data management across their organization's environment, including their core data center or centers, their edge locations, remote or branch offices, their users and employees, business partners and customers and across data that lives in public cloud or across multiple public clouds. If you collect all of this data without any understanding or organization of the data, it may make the matters even worse. Organizations need intelligent data to fuel intelligent data operations and analytics. Intelligent data operations go well beyond simple data management and governance activities and into data operations and integrated workflows to tackle tasks such as data prep, cleansing, manipulation and transformation functions to help facilitate search, analytics and evolving regulatory compliance. These newer use cases depend on robust metadata and require massive scale and high performance.

Putting Modern Workloads Into Practice

Let's double click. You have two different ways to keep track of information. Let's say you are tracking your customer purchases. One method is simple. Each sale is stored in a Microsoft Word invoice that has all of the information on the sale, and the files are saved with the order number in the filename (Order12345). The other method is to use an invoicing system (like most of us do). Suppose you receive a request to find all the orders of a particular customer. In the Word example, we open each order and note which ones belong to that customer. It takes HOURS. In the other example, we search the invoices by customer name or number and in seconds we know all their orders. We can probably also see how long they have been a customer, how much they have spent, how many orders they have placed… A host of information about the customer is available at a moment's notice. In this modern day, all your invoices are contained within the same system. But is everything? Doubtful.

Object storage has the ability to give you the same feature set as in the second example above. We can use custom metadata (metadata is data about an object and can be just about anything you can think of) and programmability add the important information to the file. Then we can search that data just like we would in a database. We can search by author, date, extension – all of which is simple. But what about application version, customer, invoice, document ID, country created, original time zone, or even really specific things like governance policy, applied date, or original application? All of these can be custom metadata fields, those and much more.

What could original application mean? How many applications do you have in your environment that exist solely because the associated data has value? You no longer use the applications, you do not access the data, but you have a regulatory requirement to keep it, so you spend thousands of dollars maintaining the application. I would bet you have 50 of those applications. Object storage means you have an alternative. Write a script to read the data from the application and write it to the object bucket, including the custom metadata that adds the context you previously relied on the application to provide. You have full query capabilities and programmatic access to the data. Making it easy to find, and even easier to incorporate into your next data application. No more old application siloes locking you out of your data.

So, now I have a place to store my data that is independent of the application used to create the data. I can enhance the data with as much custom metadata as I feel is necessary to make the data as valuable as possible to my organization. I can add more metadata at any point, as my needs and use of the data changes and evolves.

What about modern applications? Applications like Splunk. Splunk is an amazing tool that can be used to aggregate log data, create dashboards, make recommendations and help your business become more data driven. However, the more data you add to Splunk, the more complex and expensive your Splunk environment becomes. Enter object storage. Splunk data is the most valuable for the first 7 days, for the next 30 days the same data is needed but less frequently. After 3 years, the data is needed far less frequently but it remains important for trending and historical analysis. Say for example, you discover a problem that occurs every quarter. You will want to be able to go back and analyze the related data over time. Hitachi Content Platform allows you to search Splunk data, even if it has been frozen, so potential insights don't get lost forever.

What else can object storage do? You can back up data to an object store, replacing tape. If your backup application allows it, you can even have those backups done in "native file format", allowing you to provide all those enhancements we talked about previously. To take it a step further, you can programmatically read from the object store into your data lake or analytics application, which allows you to read all of that valuable data without the added load from the reads on your production applications. No application modernization is required before you start to gain value from that data.

Not Currently Running Object Storage? That Needs To Change Now

Object stores are the best tool you have to enable the future of data storage and monetize your data. Hitachi's object store, Hitachi Content Platform (HCP), is repeatedly a leader in the IDC MarketScape, Gartner Critical Capabilities and GigaOm Radar reports. It offers industry-leading reliability from a company that has been a staple of the data storage industry for decades. Additionally, while supporting traditional applications with economical and massively scalable storage, we have transformed HCP into a storage solution for high speed, Tier 1 workloads.

Modern object stores need to allow you to add all the custom metadata you need (and far more as your needs expand), and they need to be addressable programmatically (we have full Amazon S3 storage compatibility*, so not only can you write-applications to HCP using the same language as you would for the cloud), but you also need to do so without sacrificing performance requirements.

High performance allows you to take the workloads of tomorrow, those with staggering velocity, and utilize them immediately with the scale you do not know you need yet. These workloads can be anything from real time log aggregation, to IoT workloads, to transaction logs of high-performance databases. Each of those can also have custom metadata applied, further increasing their value to the organization. Think about gathering all of the logs that would otherwise go to your monitoring environment and keeping the native format for a year or more, reducing the load on the application and drastically lowering your licensing costs, or enhancing transaction records and normalizing the data so it is available for your analytics needs, on-premises or in the public cloud.

Finally, the value of object storage is diminished if you are not exploiting metadata and intelligent data management capabilities across workloads and clouds. To that end, we have an ecosystem to extend the value of HCP. Hitachi Content Platform Anywhere (HCP Anywhere) provides collaboration and sync and share to users and remote workers, as well as end-point backup to support self service restoration. HCP Anywhere Edge allows you to deliver Microsoft File Services to your users, supporting the technology and tools you already know, while enabling object storage for retention invisibly to your users (no need to re-train). My favorite, though, is Hitachi Content Intelligence, a full text index engine for HCP. Content Intelligence brings the next level of index, search, data quality, data transformation and data management to the portfolio. With Content Intelligence you can create custom workflows and take automated action. This is an amazing feature to add to your data. Set up a workflow to monitor for PPI or confidential data and apply the proper governance policy automatically. How about automatically applying tags to office documents, saving your employees hours of searching for data the old-fashioned way? How about a workflow that looks for encryption, flags the file, creates an IT ticket and locks the user out of the system in case of ransomware?

I hope you are now as excited as I am about the strategic value and business relevance of object storage. Take the next step and reach out to your Hitachi account exec and ask for a deeper look at how the Hitachi Content Platform portfolio can help. For more information visit Object Storage.

We Are Hitachi Vantara

We guide our customers from what's now to what's next by solving their digital challenges. Working alongside each customer, we apply our unmatched industrial and digital capabilities to their data and applications to benefit both business and society.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

CV-xxx-x Pace July 2020