Storage Systems Brief

The Hitachi Content Archive Platform 2.0

A New Revision and A Great Vision

Date: May, 2007

Author: Tony Asaro, Senior Analyst

Abstract: The Hitachi Content Archive Platform 2.0 is the latest revision of this best-in-class solution providing new valuable features and capabilities. This ushers in not only a more compelling solution for HDS but also a leap towards a greater vision based on the shift from structured data-dominated environments to a future in which unstructured data becomes king.

The Hitachi Content Archive Platform 2.0 (HCAP) offers a number of new capabilities that greatly enhances what ESG already considers a best-in-class platform. The HCAP solution competes with CAS solutions from EMC and NTAP. Background: HCAP is a disk-based archive used for compliance-based data and/or as a non-compliance long term archive. It is a highly intelligent object-oriented solution that supports standard protocols. HCAP can start out small and grow to a massively parallel system while remaining easy to manage regardless of size. Additionally, end-users can implement HCAP as a fully integrated solution using HDS storage or as a gateway–running the HCAP software on standard and qualified server platforms with existing heterogeneous FC or iSCSI-based storage systems. Very cool stuff.

The Hitachi Content Archive Platform 2.0 supports the following (those marked with an asterisk represent new features and capabilities):

- True clustered architecture: The ability to scale the system to meet the needs of your environment. True clustered storage provides a single logical storage system that allows any application to access any piece of data stored on it through any storage controller in the cluster. A true cluster provides for the aggregation of all its hardware resources into a single system. HCAP can scale the cluster a single node at a time or in multiples. This enables you to add more resources, including processors, bandwidth, cache memory and capacity, as needed.
- Full content and metadata search: The ability to search keywords based on metadata (name, type, custom data) and within the content itself. We believe that searching within the content is extremely valuable.
- Sophisticated retention schemes: This can be simple and straightforward and also provides you the ability to create more specific and well-defined retention policies.
- Object shredding: An important issue for security and administration.
- Support of up to 32 billion objects*: We are seeing end-users with large object counts today and forecasting those object counts to grow over the next three to five years. It is important for a digital archive solution to provide a high ceiling of object count support.
- Support of up to 20 PB of capacity*: Again, digital archives can potentially consume a great deal of disk capacity. That is why it is essential for capacity support to be able to massively scale. ESG is aware of digital archives made up of 2 PB today and the number and size of these archives continues to grow rapidly.
- Object-based replication*: This becomes essential for many environments, especially those that are heavily regulated.
- Encryption of data at rest*: ESG is seeing greater interest in encryption of data at rest within end-user companies. Again, heavily regulated industries have a great deal of pressure on them to meet compliance mandates and provide airtight data security. In fact, data privacy is becoming a major priority with end-users.
- Custom metadata*: This provides a level of control for faster and more accurate access of content. End-users can provide custom metadata that is meaningful to them in order to better utilize the information stored on HCAP.
- HiCommand integration*: This provides a single management platform to allow HDS end-users to support HCAP as well as other HDS storage systems.

- Data Liberator*: HDS provides the ability to move objects from EMC Centera systems to HCAP while maintaining the retention, immutability and metadata attributes. ESG has spoken to a number of end- users that don't want to be locked in for multiple years because of compliance regulations. Data Liberator provides these end-users with a way to move from EMC to HCAP. This is at least a start in the right direction towards complete heterogeneous data migration for CAS systems.
- Data de-duplication*: As discussed, digital archives can become quite massive very quickly. The ability to identify duplicates and only store unique data can significantly reduce the capacity requirements of your digital archive. This will lower cost, reduce floor space, impact power and cooling costs, etc. HCAP data de-duplication is very efficient, powerful and reliable.

The Hitachi Content Archive Platform 2.0 supports the following platforms (those marked with an asterisk represent new platforms):

- WMS100
- AMS200
- Disk-less appliance
- AMS500*
- AMS1000*
- USP V*
- NSC55*

The additional platform support is extremely important and valuable to end-users. First, it enables existing HDS end-users to leverage the investment they've already made in HDS storage. Additionally, it allows them to build massive archives without having to purchase a large number of HCAP controller nodes. Depending on their performance requirements, users can support a large amount of capacity with a small number of HCAP nodes. Therefore, HCAP has a flexibility advantage that end-users should appreciate. HCAP can scale to a large cluster of nodes for high performance that grows linearly and provide a large amount of capacity for environments with modest performance requirements.

HDS and the Content Archive Landscape

HDS is best known for its Enterprise-class storage systems, starting with the Lightning series up to the recently released USP V. Certainly these are great and powerful systems but the world of storage extends beyond Enterprise-class SANs. HDS appears to be enlightened to this actuality as evidenced by its actions. HDS continues to push its enterprise storage platforms but has also turned its attention to an expanding portfolio of solutions, including storage and data management software, a great NAS solution (BlueArc), an increased focus on its midrange storage systems, intelligent VTL with data de-duplication (Diligent) and perhaps the most important new solution in its portfolio– Hitachi Content Archive Platform, its content aware solution (CAS). A major indicator of the value of HCAP is that HDS did something out of character when they acquired the emerging vendor Archivas to obtain it.

The CAS market is still emerging and today is essentially a duopoly with EMC and NetApp as the dominant players and HP and IBM a distant third and fourth. Additionally, from a relative perspective, the CAS market isn't huge, with less than $1.5 Billion in overall annual revenue. However, the market potential is vast and it is still too early for anyone to declare victory. The catalyst for the CAS craze was compliance–at least in the US. Federal mandates drove public companies, financial institutions and, to some extent, healthcare providers to acquire CAS storage systems in order to retain digital data as a defense against regulators. Coming up short when regulators request files and records can have major ramifications including hefty fines or worse. CAS solutions provide Write Once Read Many (WORM) technology and, as such, can offer immutable proof that a company's data has not been altered.

Compliance is still an issue but it is no longer driving the same accelerated growth of CAS it once did. However, other far-reaching and sustainable market dynamics come into play in the CAS arena. The simple fact is that most companies and organizations create more unstructured data, including files, presentations, spreadsheets, images, graphics, etc., than any other data type. Additionally, it is becoming commonplace to create audio and video files, which consume massive amounts of capacity, even in mainstream companies. The majority of stored data is unstructured and the ongoing creation, storage, access and use of this data will drive the CAS market going forward.

Turning Data into Information

Consider the definition of "data" in this context as blocks that are stored onto disks. Today, this is a huge market, with hundreds if not thousands of applications reading and writing data onto blocked-based storage systems. The myriad of applications are each individual containers with little or no commonality between them. Further, the information that these applications create is essentially invisible or rather is unknown to the storage system or disks on which it resides. To the storage system, information is an electronic abstraction of that information, which we call data.

One of the most valuable ways that CAS transforms data into information is through federated search and indexing. The Hitachi HCAP storage system can find the content and information you're looking for by searching keywords based on filename and other metadata attributes but can also scan within documents for keywords. In this way, the quality of your searches is vastly improved and you are far more likely to find what you are actually looking for.

The name of the game is accessing the information you need quickly and precisely. Otherwise, you can easily be drowned by a sea of content. Or, conversely, you may not find the essential information that you need. It is important to have all (or as much of) your information in one place. This ensures that you are looking in the right place for what you need. It is non-productive to spend time looking for something when you are not sure whether it is there or not.

A wide variety of applications are creating data separately– there is no commonality or correlation of all of this data. To access information, you need to go to the specific applications that created it. If you are doing research on a broad topic, you may need to go to multiple applications to find all of the required information. Federated search enables leaps in efficiency since it creates commonality where none existed. It enables you to leverage all of your content independent of the various applications. HCAP provides federated search and indexing using metadata and within content, raising it from data to information storage.

The value of accessing information quickly and precisely should be self-evident. In the case of compliance, it can make the difference in saving you from a hefty fine. It could provide evidence to support your case during litigation proceedings. During a project, it could save you hours, days and perhaps weeks by providing you with the information you require. These are reactive events that are potentially critical to your business. Additionally, a content archive with quick and precise search capability can also be used proactively to help you to research new product or service ideas and create businesses.

Issues and Challenges

HDS does provide a great solution but their leading competition has a head start in the market. HDS has to be smart and aggressive about selling the HCAP solution. The short-term potential is for them to win some big deals and create a new revenue stream. The long-term has much greater implications--being a leader in the new storage landscape in which unstructured content dominates.

ESG's View

The Hitachi Content Archive Platform provides one of the best content archive solutions out there, from multiple perspectives, including intelligence, scalability, performance, flexibility and ease-of-management. Version 2.0 of HCAP is feature-rich, providing a compelling offering in the HDS arsenal, giving them real potential to be a leader in this space. The IT world is moving from the structured data (database) archetype to an environment dominated by unstructured data (files, images, multi-media). HDS is one of the kings of structured data storage and is poised to be a leader in the next wave.