Big Data – Shining the Light on Enterprise Dark Data

WebTech Q&A Session, April 17, 2013

1. **Metadata can be overwhelming. How does Hitachi Content Platform (HCP) help control metadata?**
   In several ways. HCP:
   - Adds a structured schema to metadata and can discover hierarchical relationships between tags
   - Offers massive parallelism and the ability to rapidly ingest a large volume of metadata
   - Has a built-in metadata query API to make analytics easier
   - Allows many applications to have distinct custom metadata annotations, and can tag a piece of content in many different ways
   - Allows you to be selective in how much metadata you index

2. **Does the Hitachi solution support NetApp, Windows, or other platforms?**
   The indexing and searching component, Hitachi Data Discovery Suite (HDDS), is tightly integrated with Hitachi solutions and can also support NetApp, Windows, and any CIFS or NFS platform.

3. **What kind of system do I need to get all this done?**
   You can start small and add more nodes without impacting an existing implementation. For the search solution, which is an appliance (hardware and software), you can start with one node and add additional nodes as your environment grows. If you add an additional node, HCP automatically and evenly distributes the indexes.

   With HCP, you can start out with a system as small as 4 TB and you can scale up to 40 PB. HCP scales out in one of two ways:
   - Adding nodes to increase processing power (faster ingesting and indexing of data)
   - Adding disks to increase storage space

4. **How do you create metadata when an application will not send that data into the archive?**
   In Hitachi Content Platform v6, we made it easier for applications to send metadata to HCP with a feature that can discover the structure of metadata by examining the data held by an application.

   We have a prototype of a tool that does a great job of extracting metadata from data. Some customers have already benefitted from that tool, which we expect to include in an upcoming release.