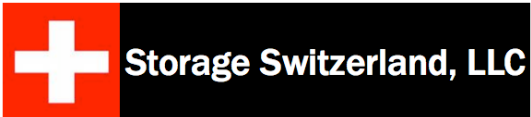


STORAGE SWITZERLAND REPORT

INDEX ENGINES 3.0 - DATA DISCOVERY FOR IT

Thursday, October 1, 2009



George Crump, Senior Analyst

[Index Engines](#) has long been a leader in the content indexing and discovery marketplace. Their focus, up until the recent 3.0 release, has been primarily on discovery for litigation readiness with the unique ability to scan backup tapes, enabling rapid search and extraction of the content without the original backup software. This solution has enabled corporate clients to proactively understand what is on the mountains of tapes they have accumulated, and then determine if they need to archive specific content or simply recycle the tape.

In 3.0 Index Engines combines the indexing of primary storage data with backup data that previously required separate processing. Index Engines 3.0 brings a unified view to data indexed from network storage and backup tapes and allows one point access for data discovery and management. This streamlined approach enables IT to leverage eDiscovery efforts to meet larger 'IT Discovery' goals.

In a recent announcement with BlueArc, Index Engines described an implementation where they validated the indexing of millions of unstructured files and email messages at the rate of 1TB per hour with a single indexing appliance. This is a dramatic increase in performance, as other data indexing platforms require 4 or 5 appliances to achieve their tops speeds and still don't approach 1 TB/hr. This speed makes indexing a more practical tool for day-

to-day use in the enterprise. With one appliance, the environment can be comfortably indexed every day, and those results reliably used by IT to help facilitate the ongoing need to 'find and mind' information that is hidden on servers.

IT Discovery differs from eDiscovery in the sheer magnitude of data that it deals with. This means discovery tools built for IT must handle much more data and must be oriented towards performance and capacity. When these two requirements are met, you get a system that can index bulk enterprise data in hours and use as few appliances as possible. Prior to this, if there was a desire for data-center-wide indexing, the IT staff would have to deploy dozens of appliances to index the entire data set, and the project would span months.

Source of Speed

Where does this speed come from? First, it is important to know that most indexing tools use a variety of third party and open source modules originally designed for internet applications. Your data center is not the internet. Index Engines built their index module from the ground up to solve the indexing needs of the enterprise. All components of the module, including the database, word scraping and query engine have been developed by Index Engines for this sole task.

Second, the data is scanned and processed sequentially in order to maintain high-speed throughput. Without sequential processing, a part of Index Engines intellectual property, other indexing solutions require random disk access which severely handicaps performance and makes the solution susceptible to hardware limitations.

Finally, Index Engines leverages sequential protocols for data delivery. For example to process unstructured data on file systems, Index Engines when applicable, uses NDMP to scan data as opposed to the traditional use of CIFS and NFS. NDMP is significantly faster at moving large chunks of data over a network. While CIFS and NFS are supported, NDMP delivers the most efficient performance.

For email data Index Engines goes beyond the normal access methods and has done the hard work of directly understanding the format. For example in Exchange environments, most eDiscovery appliances use MAPI to access the information. Just like in backup, accessing data via this protocol causes a significant performance issue. Other indexing appliances are quick to blame Exchange. Index Engines on the other hand has technology that takes a bottom up approach and performs bit for bit processing of the entire database and then leverages its sequential processing technology for fast, consistent access.

Beyond Speed

In addition to the speed of the solution the other key to the enterprise is the efficiency of the database. eDiscovery products typically have an index database whose size is 20% to 110% of the indexed data. The last thing storage managers want is to increase their storage management burden. Adding another database that's 20% (or more) of the existing capacity of the enterprise will create significant

additional load. By creating an indexing-specific database from the ground up, the Index Engines index footprint is typically 4 to 8% of the original data source. As a result of this significantly compressed index footprint, the Index Engines platform can scale to handle up to a billion files and email in one appliance. This again is far beyond other solutions and is due to the purpose-built nature of the technology that was architected with the enterprise in mind.

Storage Switzerland Take

eDiscovery has only been justifiable for a corner case of the data center; litigation readiness. Yet the core capability, to find exactly what you need exactly when you need it, has value for the entire enterprise. The limiting factor has been the resources and expense required to capture and index all the information for the entire enterprise. Index Engines seems to address this key limiter by delivering a platform that rapidly ingests large amounts of information. This transforms eDiscovery into IT Discovery and as a result delivers a significant pay off to those organizations that choose to implement it.

Bottom Line

With these advances of single node, high performance, sequential indexing, the process is ready to come out of the litigation corner and into IT Mainstream. This will allow users and IT personnel to more effectively use the data assets that they have stored in their enterprises. IT Discovery gives IT the ability to know what to save, what to move, what to archive, and what to delete. By doing all this in almost a real time fashion, suddenly data becomes an asset.

About Storage Switzerland

Storage Switzerland, is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: <http://www.storage-switzerland.com>.