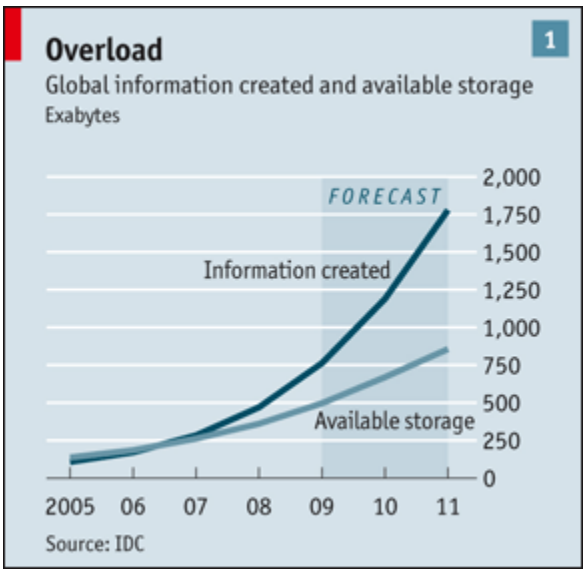
Giovanna Corridoni

Flash Research Assignment: Servers and Storage Technologies

Data is growing at an exponentially fast rate that is surpassing the storage capabilities available today. We need to invest in a technology that enables us to maximize storage utilization for the vast amount of data our company houses. Data deduplication is a consolidation process that will allow us to remove redundant data from our system decreasing operating costs significantly.

Data deduplication is a form of consolidation used to remove redundant data in our systems to maximize storage utilization. The problem with large amounts of data is not that there is so much, but rather that so much is being repeated. The replicated data that usually comes from system backups, emails, and documents shared by multiple users is unnecessary and causes inefficiencies in our storage disks. However, data deduplication, or intelligent data compression, uses a specific algorithm to detect and eliminate redundant data created in systems. The process begins by sorting through data to identify unique chunks of data, which are then stored leaving only leaving a single copy of the unmatched data. This consolidation process runs frequently on an automated basis. With data deduplication running multiple times over the same data, the algorithm compares data that is currently being analyzed with the stored copy; whenever redundant data is detected, it is replaced with a small reference to point to the stored chunk, leaving only one copy of the same data. Furthermore, when there is an alteration in a dataset or new data is added, an original copy is then stored onto the storage disk. Data deduplication therefore eliminates repetitious information, creating more readily useful space on each physical storage device.

Investing in data deduplication will allow our business to fully utilize our storage devices to the maximum potential and therefore save money and physical space on each storage device. Removing redundant data means more space on each storage disk and less disks. With fewer disks, we can reduce our operating costs by spending less on purchasing, deploying, powering and cooling each storage disks. Additionally with consolidated data on our disks, we can store data longer, which means spending less money on slower backup and recovery technologies, such as magnetic tape, because the data stays in-house longer. The benefits of using data deduplication will be realized in operation costs and system performance.



http://www.economist.com/node/15557443?story_id=15557443

Bibliography

"Data, data everywhere | The Economist." *The Economist - World News, Politics, Economics, Business & Finance*. N.p., 25 Feb. 2010. Web. 27 Feb. 2013. <http://www.economist.com/node/15557443?story_id=15557443>.

N. Mandagere, P. Zhou, M. A. Smith, e S. Uttamchandani, «Demystifying data deduplication», in *Proceedings of the ACM/IFIP/USENIX Middleware '08 Conference Companion*, New York, NY, USA, 2008, p. 12–17.

Russel, Dave. "Hype Cycle for Storage Technologies, 2012." *Technology Research | Gartner Inc.* Gartner Inc., 5 July 2012. Web. 26 Feb. 2013.