1. **Establish governance or choke the benefits of the data storage evolution**

Blinzler, Michael W.  (Science in Operations Management, University of Arkansas, United States)

Source: Annual International Conference of the American Society for Engineering Management 2011, ASEM 2011, p 271-280,  2011,  Annual International Conference of the American Society for Engineering Management 2011, ASEM 2011

**Abstract:**

The data storage evolution is underway as business units and departments capitalize on rapid improvements in capabilities and declines in costs. Companies must now establish appropriate data governance or benefits will be choked by clutter and compromised data. This data storage evolution is reviewed holistically through the lens of current needs and opportunities. Cloud computing, virtualization, file shares, Microsoft SharePoint and Google Apps, email, backups, COLD storage, and OCR (Optical Code Recognition) technologies are analyzed. The exhaustive collection of what has been learned about data governance, including records management, is mined to determine what is possible. Taxonomy, data classification, e-discovery, and legal requirements are considered. A SWOT (Strengths, Weaknesses, Opportunities, and Threats) is used to integrate data governance and the data storage evolution. Strengths and opportunities are used to compensate for weaknesses and threats. A project plan that addresses both social and technical aspects is provided as a path forward. Copyright, American Society for Engineering Management, 2011.(30 refs)

1. **Efficiently binding data to owners in distributed content-addressable storage systems**

Eaton, Patrick (University of California, Berkeley); Weatherspoon, Hakim; Kubiatowicz, John

Source: Proceedings - Third IEEE International Security in Storage Workshop, SISW 2005,  v 2005, p 40-51,  2005,  Proceedings - Third IEEE International Security in Storage Workshop, SISW 2005

Abstract:

Distributed content-addressable storage systems use self-verifying data to protect data integrity and to enable graceful scaling. One feature commonly missing from these systems, however, is the ability to identify the owner of a piece of data in a non-repudiable manner. While a solution that associates a certificate with each block of data is conceptually simple, researchers have traditionally claimed that the cost of creating and maintaining certificates is too great. In this paper, we demonstrate that systems can, in fact, efficiently map data to its owner in a secure and non-repudiable fashion. To reduce the cost of creating and maintaining certificates, we extend the traditional content-addressable interface to allow the aggregation of many small data blocks into larger containers. The aggregation is performed in a way that also supports self-verifying data at the granularity of the block and container, fine-granularity access, and incremental updates. We describe two proto-type implementations and present preliminary performance results from deployments on PlanetLab and a local cluster. © 2005 IEEE.(16 refs)

1. **On-line storage management with distributed decision making for content-centric networks**

Sourlas, Vasilis (Department of Computer and Communication Engineering, University of Thessaly, Greece); Gkatzikis, Lazaros; Tassiulas, Leandros Source: 2011 7th EURO-NGI Conference on Next Generation Internet Networks, NGI 2011 - Proceedings,  2011,  2011 7th EURO-NGI Conference on Next Generation Internet Networks, NGI 2011 – Proceedings

**Abstract:**

Content traffic proliferation in Internet makes more dire than ever the development of radical new network architectures, where information will be addressed by semantic attributes rather than the origin and destination identities. In this direction, content-centric networking appears as a flexible communication model that meets the requirements of the content distribution trends of the future Internet. In such networks, information will reside at various locations/nodes (the Content Delivery Network surrogate servers) and the requests of the users for some piece of information will be directed to the closest replica. Since the location of the users and the popularity of the content varies over time, the problem of finding the optimal replication pattern for the available content, given the storage constraints, comes into the foreground. In this paper, we propose two on-line storage management algorithms of gradient descent type, designed specifically for content-centric networks. The proposed algorithms are of polynomial complexity and thus adapt easily to any environmental changes. Each node re-assigns its information items with the aim to minimize the overall traffic cost of the content delivery as the popularity and locality of users' requests change. While both the proposed algorithms operate in a distributed way, differ in the amount of information required for the decision making. Thus, we identify the inherent information performance tradeoff and compare them in terms of network traffic, convergence speed and amount of circulated information. © 2011 IEEE.(27 refs)

1. **Creating a 'Cloud Storage' mashup for high performance, low cost content delivery**

Broberg, James (Department of Computer Science and Software Engineering, University of Melbourne, Australia); Buyya, Rajkumar; Tari, Zahir Source: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),  v 5472 LNCS, p 178-183,  2009,  Service-Oriented Computing - ICSOC 2008 Workshops - ICSOC 2008 International Workshops, Revised Selected Papers

Abstract:

Many '**Cloud** **Storage**' providers have launched in the last two years, providing internet accessible data **storage** and delivery in several continents that is backed by rigourous Service Level Agreements (SLAs), guaranteeing specific performance and uptime targets. The facilities offered by these providers is leveraged by developers via provider-specific Web Service APIs. For content creators, these providers have emerged as a genuine alternative to dedicated Content Delivery Networks (CDNs) for global file **storage** and delivery, as they are significantly cheaper, have comparable performance and no ongoing contract obligations. As a result, the idea of utilising **Storage** **Clouds** as a 'poor mans' CDN is very enticing. However, many of these '**Cloud** **Storage**' providers are merely basic **storage** services, and do not offer the capabilities of a fully-featured CDN such as intelligent replication, failover, load redirection and load balancing. Furthermore, they can be difficult to use for non-developers, as each service is best utilised via unique web services or programmer APIs. In this paper we describe the design, architecture, implementation and user-experience of MetaCDN, a system that integrates these '**Cloud** **Storage**' providers into an unified CDN service that provides high performance, low cost, geographically distributed content **storage** and delivery for content creators, and is managed by an easy to use web portal. © 2009 Springer Berlin Heidelberg.(8 refs)

1. **Structured and unstructured data retention solution**

Rajecki, Keith Richard  (Industry Solutions, Oracle Corporation, Redwood Shores, CA, United States)

Source: Archiving 2014 - Final Program and Proceedings, p 40-41,  2014,  Archiving 2014 - Final Program and Proceedings

Abstract:

**Data** **Retention** solutions are increasingly important as the world becomes more digital. These solutions help museums, national libraries, governments, educational institutions, and corporations world-wide capture and manage institutional assets. **Data** **Retention** provides a means of organizing institutional **data** while providing greater levels of access and controls. Today's **Data** **Retention** solutions must be extensible to meet the evolving business needs of Big **Data** and Cloud Computing. While Hierarchical Storage Solutions and tiered storage provide the most reliable and cost effective solution, access and performance requirements are also important design factors. If architected incorrectly, **data** searching and retrieval can be time consuming and costly. Traditional tape-only archival methods simply do not meet the access requirements of many of today's repositories and long-term archives. Likewise, storing all the **data** on disk requires greater administration and is more costly. The proposed **Data** **Retention** solution provides a proven solution with a balance between disk, tape, and cloud storage to support long term archiving of **structured** and **unstructured** **data**. © 2014 Society for Imaging Science and Technology

1. **Unstructured data management shifts industry to unified IT infrastructure.**

By: Terranova, Manuel.

Offshore. Feb2015, Vol. 75 Issue 2, p58-61. 2p.

Abstract: The article focuses on the management of unstructured data which has changed the offshore industry into an information technology (IT) infrastructure and investment. Topics include the growth of raw data sizes and advances of seismic data acquisition technology, the concerns on the raw data's unstructured nature during a seismic survey, and the need for companies to re-architect the computing infrastructure.

<http://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=101065136&site=ehost-live>

1. **Key Considerations For Managing Unstructured Data.**

By: Ali, Shahbaz.

Pipeline & Gas Journal. Jul2013, Vol. 240 Issue 7, p46-47. 2p.

Abstract: The article provides information regarding unstructured data management. It states that the three pillars for unstructured data management include storing unstructured data, risk controlling, and comprehending information. It notes that data protection is relevant in the energy industry. It mentions that data management platforms must handle the development of discovery data from upstream processes, and deliver data retention, protection, and high availability access.

<http://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=89467332&site=ehost-live>

6. **Answering Big Data’s 10 Biggest Vision and Strategy Questions**

<https://www.gartner.com/doc/2822220/answering-big-datas--biggest?docdisp=share&srcId=1-4398736771>