



Title: Understanding Your Data Helps to Reduce Data Storage Costs.

Author: Bill Mottram

Date: November, 2008

Abstract: *While budget growth is lethargic at best, demands for on-line data storage continues to grow and the drivers of power consumption are becoming visible at senior executive levels. Status Quo will not be acceptable and hence the need to re-tool the thinking surrounding the management and storage of data.*

Data classification enables the storage administrator to understand the expectations associated with the corporate data under their stewardship. With this understanding, inactive data can be migrated to a less expensive class of storage freeing up expensive tier 1 storage to deliver the performance required for highly dynamic, business critical, Tier 1 data. In short matching the user expectations for data to its cost of storage.

In most organizations data is not growing, it is exploding! The problem with data is that once it is created it is rarely eliminated whether by choice (corporate governance), indolence (poor data management practices) or compliance (government legislation). With the accretive nature that characterizes data growth the problem is only going to become more acute, particularly if current storage practices continue.

All data, whether transactional or persistent (fixed content), may have the same basic binary characteristic but not the same usage patterns. Particularly in today's unstructured, content rich, image and video based, WEB 2.0 world. The criticality of content and its usage pattern establishes the primary data characteristics that differentiates data and enables storage useful classification. Different data has different performance expectations with access latencies bridging from milliseconds to seconds or even days (how about never),

Examples of compliance driven retention periods.

■ tax returns -	Indefinitely
■ expense reports -	6 yrs
■ bank statements -	3 yrs
■ production correspondence -	8 yrs
■ terminated personnel files-	6 yrs
■ employees exposed to toxic substances -	30 years
■ medical records -	Life + 2 yrs
■ monthly trial balances -	6 yrs
■ insurance claims post settlement	10 yrs
■ manifests -	4 yrs

depending on the data in question. Data characteristics are defined by access requirements, retention time, integrity requirements, content and technology longevity, all variables that impact how data should be managed, stored or deleted and on what storage technology.

So why should data characteristics drive storage selection? Simply put, knowingly placing inactive data on a tier 1 platform is just as daft as expecting tier 3 or 4 storage to satisfy your service levels obligations for business critical, transactional data. Neither scenario illustrates good stewardship of corporate assets whether digital or physical. While scenario two is unlikely, unfortunately, inactive data is consuming significant tier 1 resource in many data centers. According to industry analysts such as ESG and the Tanaja Group the estimate is that inactive or persistent data represents 70% to 80% of data within the data center¹. Interestingly in a recent survey of storage professionals conducted by COPAN Systems, 55% of respondents thought that this number was under 50% with 20% who did not know how much persistent data they had never mind managing it. Of those who were aware of their persistent data 31% used primary storage and 40% used a combination of archive and primary storage to store and manage persistent data. Yes the persistent data was consuming expensive storage, but it was safe and available if needed and by the way the IT Manager was not accountable for the utility cost of these gluttonous, power hungry storage frames. However, realities are changing and with budget growth being lethargic at best, most likely negative, more prudent management of storage resources is required.

Data makes up the digital history of an enterprise. It is a corporate asset which holds considerable value. However, the value of data is variable and tends to be time and activity dependent. Value can be influenced by the age of the data, by access requests whether internal or external, and while not characteristics per se its *find-ability*, its time to be accessed and its integrity are all critical influencers. With that being said, there is also much data being stored and managed in an enterprise that is worthless and its retention and management is a measurable expense to the enterprise. Hence, the need for robust data classification and management practices.

Data types can be simply classified as either highly active, transactional data or inactive, primarily historical/reference data, becoming known as persistent data.

1. *Transactional Data* – this is the tradition view of data and is the view that has molded today’s disk storage architectures. It tends to be data that is being captured or created, is highly dynamic in nature, drives high IOP, is random in nature and tends to have a short shelf life. This is why traditional, transactional designs are a read-write-modify access model, they are optimized to provide access to data at all times, they are cacheable or demonstrate temporal and spatial locality and are optimized to small grain data access.
2. *Persistent Data* – this is data that once created is rarely accessed or rarely if ever modified. This is data that does not demand the same response time, low IOP demand and tends to have a low temporal access locality meaning caching is a wasted expense. Persistent Data tends to have a long term retention requirement, is bandwidth centric, has data integrity concerns, likely to be event driven, immutable, reference content and is the fastest growing segment of today’s digital information. As referenced earlier 70% to 80% of data in a data center fit this “Persistent Data” description.

¹ Independent surveys from the Enterprise Storage Group and the Tanaja Group indicate that 70% to 80% of data in a data center in persistent.

Differentiating data types does not enhance or diminish the relative importance or value of data, it simply improves the chances for its cost efficient storage, its effective management, availability and use. There is agreement that persistent data is the fastest growing data type, in terms of data volume in the data center. The reason is that much of this data is subjected to minimum retention periods that are dictated by one compliance regulation or another. Not only must this data be retained but when requested it must be available in a timely manner. There is a history of significant financial penalties levied on companies who failed to deliver data in the time required by law.

By appreciating that different data has different value and be willing to execute even on the simplest of classifications enables the astute data storage manager to effectively match the value and requirements of their data to the cost and performance of the “hosting” storage technology.

This realization is the epiphany that opens the door of opportunity to significant cost savings while surviving today’s phenomena of explosive data growth.

Bio: Bill Mottram compliments his role the managing partner of Verdictus Associates Inc, a high technology marketing consultancy as analyst practitioner with Data Mobility. He has over 25 years product development, marketing and sales experience in the data storage and information technology industries. Experienced with Fortune 500 companies such as StorageTek, Compaq and HP and smaller, start-up enterprises including Pillar Data Systems and COPAN Systems he was the marketing executive and leader responsible for the development , introduction and “go-to-market activities for a number of innovative and highly successful data storage solutions. Although primarily a marketing professional Mottram has considerable expertise in technology management, corporate acquisition and public company funding.

A graduate of Paisley College of Technology (electronic engineering) and holds a Masters of Business Administration from Pepperdine University. Contact Mottram at:

email: bill@verdictusassociates.com WEB: www.verdictusassociates.com
Blog: www.storagetopics.com

Copyright © 2008 Verdictus Associates Inc. All Rights Reserved. Reproduction of this publication without prior written permission is forbidden. Verdictus Associates Inc believes the statements contained herein are based on accurate and reliable information. However, because information is provided to Verdictus Associates Inc from various sources, we cannot warrant that this publication is complete and error-free. Verdictus Associates Inc disclaims all implied warranties, including warranties of merchantability or fitness for a particular purpose. Verdictus Associates Inc shall have no liability for any direct, incidental, special, or consequential damages or lost profits. The opinions expressed herein are subject to change without notice.