



Surviving the *Perfect Storm* in Data Management

Dave Clements
FileTek Business Development
dclements@filetek.com

Publication Number:
Z24-V1.0 Rev. A.

August 1, 2000

The FileTek logo consists of the word "FileTek" in white text on a teal square background.

FileTek

All rights reserved. No part of this publication may be reproduced, translated, stored in any electronic retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of FileTek, Inc.

© 2000 by FileTek, Inc., Rockville, MD
Publication Number Z24-V1.0, Rev. A
August 1, 2000

FileTek and StorHouse are registered U.S. trademarks of FileTek, Inc. All other brand or product names are trademarks or registered trademarks of their respective owners. Documentation for FileTek's StorHouse product. Protected by the following U.S. patents: 4,864,572; 5,247,660; 5,727,197; 6,049,804. Other patents pending.

Abstract

The emergence of e-applications is creating extremely high volumes of new data. Indeed, the collection of information that analysts so calmly referred to as a “sea of data” just ten years ago has now swollen to Tsunami forces. As a result, many analysts might say that today’s data management professionals are facing a *Perfect Storm*. If they don’t find calmer waters soon, they’ll drown.

This paper is for information technology and data management executives who (1) understand the value of maintaining actual event records rather than their summaries, but (2) have not yet determined a cost-effective data management strategy for doing so with large and growing volumes of data. It introduces the concepts of *active* versus *supportive* data management needs, *high-speed* versus *high-volume* storage technology requirements, and *micro-* versus *macro-analytics*. In addition, the paper recommends new *data management infrastructure* alternatives and provides their associated cost reductions.

By adopting the recommendations in this paper, planners can strategically look for, consider, and implement informational initiatives and concepts that, in the past, they may have immediately discarded due to prohibitive costs.

The Wave: Explosive Data Growth

Most analytical reports today cite exorbitant factors when projecting enterprise data growth over the next five years. For example:

- *Red Herring’s* March 2000 summary, “The Age of Petabytes,” forecasted data growth rates of 75% to 150% per year.
- A META Group analyst who spoke at a May 2000 eCRM conference projected data increases of a hundredfold within five years through the year 2004. Enterprises that are having difficulty coping with three terabytes (TB) of data today need to quickly find solutions for dealing with 300 terabytes of data tomorrow. Since the May 2000 conference, other META representatives have validated this growth factor as well as the urgent need for in-depth strategic data management planning.

- A recent Deutsche Banc Alex.Brown data study¹ found that e-business data will grow from 30% of the total data in the first year of activity (1999 in most cases) to 75% of the total data in the fourth year. This growth represents a commanding data swell of 400% per year.

Figure 1 integrates the data in these three studies to arrive at a seemingly dependable consensus about the rate of data growth. It uses the Red Herring data as the basis of the graph. The triangulated outlook applies to Global 2000 enterprises and assumes an average starting point of three terabytes total in-house data in 1999. The striped sections of the color-coded vertical bars estimate the percentage of growth stimulated and consumed by e-business activities.

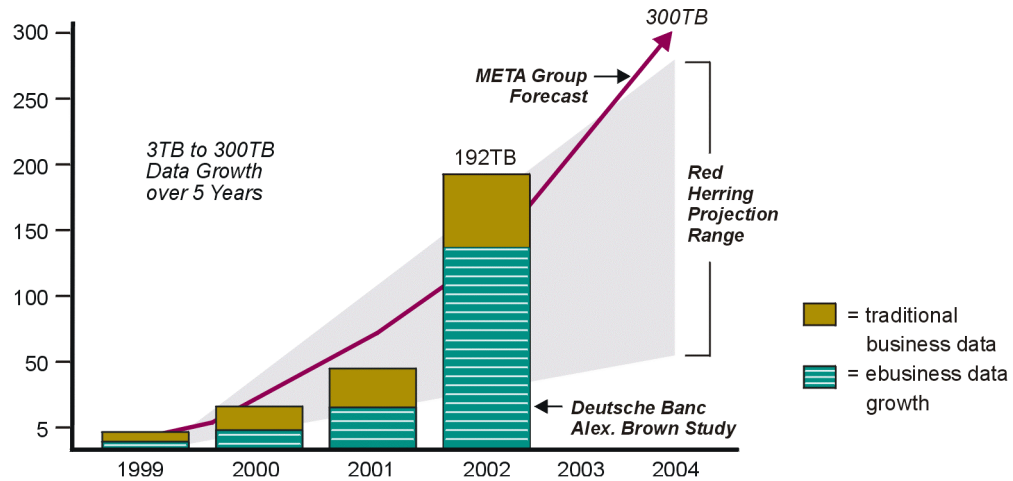


Figure 1: Data Growth Projections

Understanding and acceptance of these predictions comes only after you consider the scope of new business initiatives and the technological capabilities that both enable and support them. New e-business applications like web-front management (clickstream analysis), one-to-one customer relationship management (CRM), personalization and encounter management, supply chain management, call event detail analysis, and digital certification significantly add to an enterprise's existing IT-supported agenda. In addition, new non-scalar data types (objects) including images (drawings, X-rays, etc.), streaming audio, and video dramatically expand the data inventory.

Surging TCO Costs

It's easy to suppose that somehow more disk storage and technological advances will ease the cost of managing this tidal wave of new data. However, as Figure 2 indicates, even with the expected continuation in RAID price/TB decline, the total expenditure necessary to accommodate the projected data growth will escalate more than tenfold in the next five years.

To be consistent with the data in Figure 1, Figure 2 uses the following numbers:

- A 1999 total storage starting point of three terabytes of serviced data, which grows at a rate of 150% per year
- \$300K/TB as the starting cost of disk storage
- A decline in storage costs of 30% per year, which is the figure projected by most analysts, including IDC, Gartner, and META
- An expenditure calculation that uses a total cost of ownership (TCO) composite, which takes into account the hardware price/TB plus overhead factors for storage and data management tools and services.

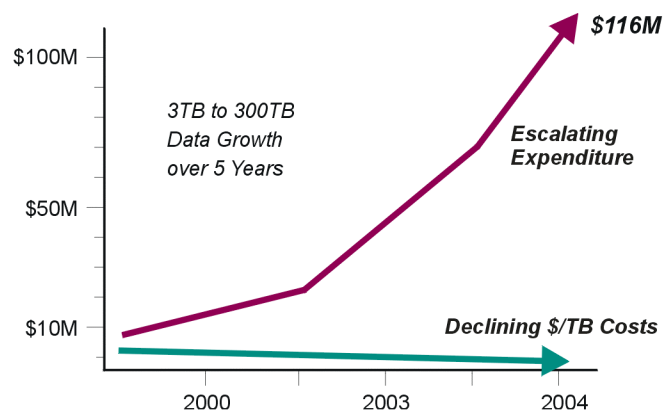


Figure 2: Storage Pricing and Expenditure

The message expressed in Figure 2 is startling. Over the next five years, given a sixfold decrease in price/TB and a hundredfold increase in data, you can expect a thirteenfold increase in total data management costs.

Alternatives: Finding New Harbors

Most IT professionals would agree that these soaring expenditures are unacceptable. Although the promise of web-enabled business applications offers greater profitability, conscientious CIOs must look for, consider, and embrace more cost-effective data management strategies and alternatives.

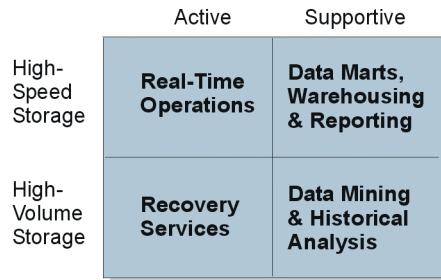


Figure 3: Data Management Framework

One of the first things to consider (or reconsider) is data management treatment. How do you use and effectively manage different types of data?

Figure 3 classifies data management treatment into a framework with four distinct categories of information processing. Each category requires different software and technology.

First, there's active versus supportive data. You must keep active data fresh and current because it's used by operational procedures. In contrast, you may derive supportive data and refresh it periodically (for example, nightly or weekly). Informational and analytical applications use supportive data.

Next, there's high-speed versus high-volume storage. High-speed storage typically consists of rotating disk memories that allow access to small amounts of data in milliseconds. High-volume storage, such as FileTek's StorHouse® data management solution, typically deploys multiple media (disk, optical and tape) and allows very large amounts of data to be accessed in seconds to minutes, depending on the configuration.

In most cases, the majority of enterprise data does not need to be maintained in an active profile on high-speed storage media. As Figure 4 indicates, as little as 15% of the total data resource may be all that is required by real-time operations, depending on their applications.

Aging details like clickstream logs, order entry line items, call event detail records, inactive account

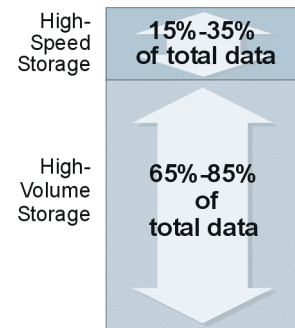


Figure 4: Data Allocation Framework

descriptions, audit summary backups, and many other file segments that do require sub-second retrieval rates can be allocated to more cost-effective, high-volume storage.

An Analytics Alternative: Charting the Course

Analytics have always been key to successful business endeavors. However, in the e-business arena, the time element for many applications is so critical that enterprises can no longer afford the luxury of offline decision making. Many of the manual processes that previously depended on decision support systems (DSS) must now be automated.

META Group's infrastructure recommendations at the May 2000 eCRM conference further endorse the active versus supportive, high-speed versus high-volume storage framework. For example, META suggests that eCRM systems should be bolstered with two levels of analytics: real-time (active) and batch (supportive). META refers to this split as micro- and macro-analytics, respectively. (Refer to Figure 5.)

Within the batch portion of the storage framework, long-running macro data mining applications culture attributes and triggers that are posted in an account master record in an on-line operational data store (ODS). In turn, the real-time micro-analytics run on top of the ODS like a state machine, and react in accordance with the derived triggers and attributes. For example, a reaction to a stored trigger might be a notification message about new product offerings that parallel a customer's previous buying trends.

This delegation of process has two areas of positive impact. Shifting the macro-analytics to the supportive/high-volume data management environment improves the real-time performance factor. Furthermore, transferring the high-volume detail data to a more economical storage platform realizes tremendous cost savings.

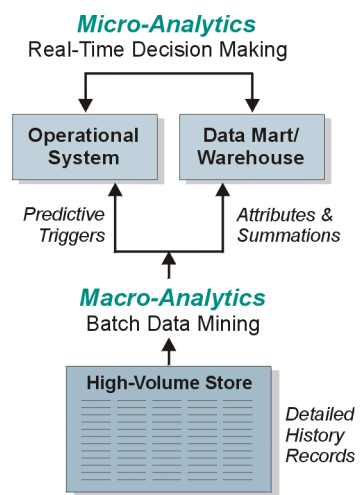


Figure 5: Two-Tier Analytics

The Economical Result: Calmer Waters

Figure 6 depicts a vastly improved cost projection based on using the proposed high-volume/high-speed data allocation framework and the two-tiered analytical model. These projections include the integration of FileTek’s StorHouse data management solution. StorHouse satisfies the high-volume requirements of the proposed data management framework in Figure 3 by providing data recovery and deep volumes of atomic detail data for supportive analytics, all in one cohesive strategy.

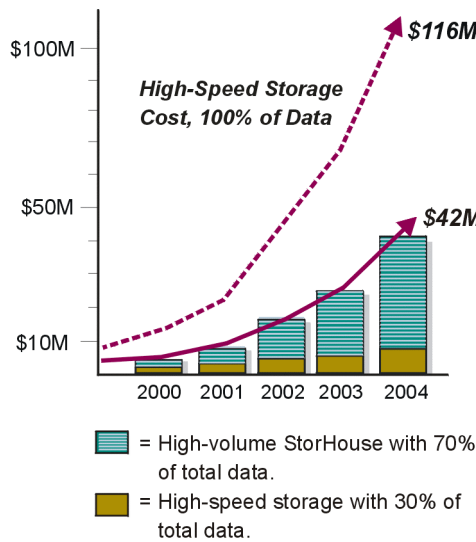


Figure 6: An Economical Alternative

StorHouse accommodates high-volume needs with RAID media and cost-effective robotic storage (automated seek-and-play elements that include optical disk jukeboxes and high-speed automated cartridge tape libraries). It also provides sophisticated storage management and relational database management software. StorHouse/SM, the storage management component, automates critical system management tasks like data migration, backup, and recovery. StorHouse/RM, the relational management component, provides row/record level selectivity regardless of database size or location of data within the storage hierarchy.

In Figure 6, the broken line represents the TCO mapping initially described in Figure 2. Contrast this line with the actual cost of storing high-volume data on a more appropriate platform. As you can see, by taking the same volume of data—but now delegating it to the proposed subsystem alternatives—you can achieve tremendous cost savings.

The example in Figure 6 keeps 30% of the data on more costly high-response RAID technology and migrates 70% of the data to a high-volume technology such as StorHouse.

Although this scenario is somewhat conservative, it represents a savings of over \$160 million over the charted five-year period. Every additional 1% of data shifted from high-speed storage to StorHouse by year 2004 will represent even more savings.

Summary

Much like nature's *Perfect Storm*, several forces in the IT world are converging to form a magnitude of data management problems that transcend previous levels of experience. Both the supply and demand sides of the business information equation are escalating together at a whirlwind pace. New-generation eApplications are producing torrents of data at a predicted hundredfold, five-year growth rate. At the same time, enterprise managers continue to thirst for more insights that can only be gained from analyzing these massive amounts of accurate and timely detail data.

Add to this turbulence an overwhelming projection in the associated cost to contain and manage these torrents of data, and it becomes apparent that new alternatives, such as StorHouse, need to be considered. For more information about StorHouse solutions, visit the FileTek web site at www.filetek.com or call FileTek headquarters at (301) 251-0600.

¹ Dolan, Timothy J., C.F.A., "eCRM: The Difference Between Winners and Losers in the e-Business World of the 21st Century." Deutsche Banc Alex. Brown, North American Equity Research/US Enterprise Software, September 15, 1999.