The Growing Importance of Archive

Prepared for **StorageTek**

By **Robert C. Abraham**
Freeman Reports

April 2005

Executive Summary

Computer storage devices are used for several purposes, such as storing **primary** computer data, acting as **data protection** devices for the primary data set, and retaining long-term **archive** data. Archived data accounts for more than 80% of all stored data and is increasingly more important to the owner of the data.

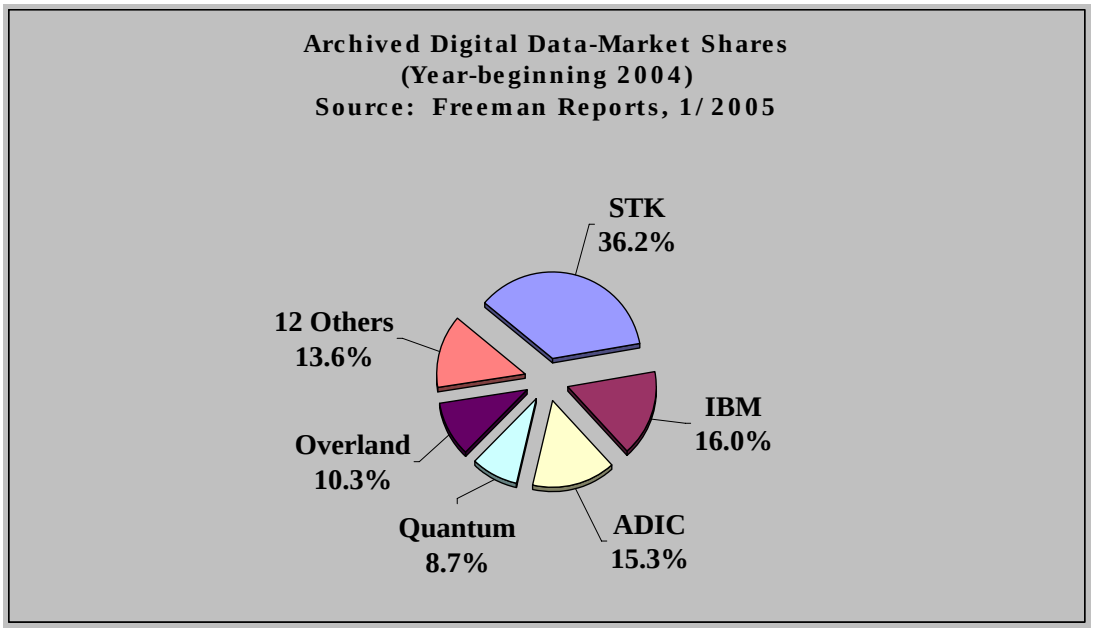
Archiving is the process of saving and preserving data files from primary or secondary storage onto a storage media to be retained for a reasonably long period of time (contrasted with short-term data protection such as backup, mirroring, replication, and snapshot). Records retention requirements resulting from compliance regulations are driving rapid growth in archiving. The massive quantity of data and the rapid growth of the installed base of archived data demand cost effective solutions.

The criteria for selecting an archive solution should include the following factors: **write-once capability, stability of long-term data integrity, restore requirements, ease of media removability and transportability, and total cost of ownership.**

Freeman Reports recently completed a major study of archiving. The purpose of the study was to determine the size and growth opportunity of the archive market and to examine the role that StorageTek plays in this space. Freeman Reports' extensive historic database was used to determine key market characteristics.

Summary of Findings

- ✓ **StorageTek is the Vendor of Choice for Archiving, with more than twice the market share of its nearest competitor.**



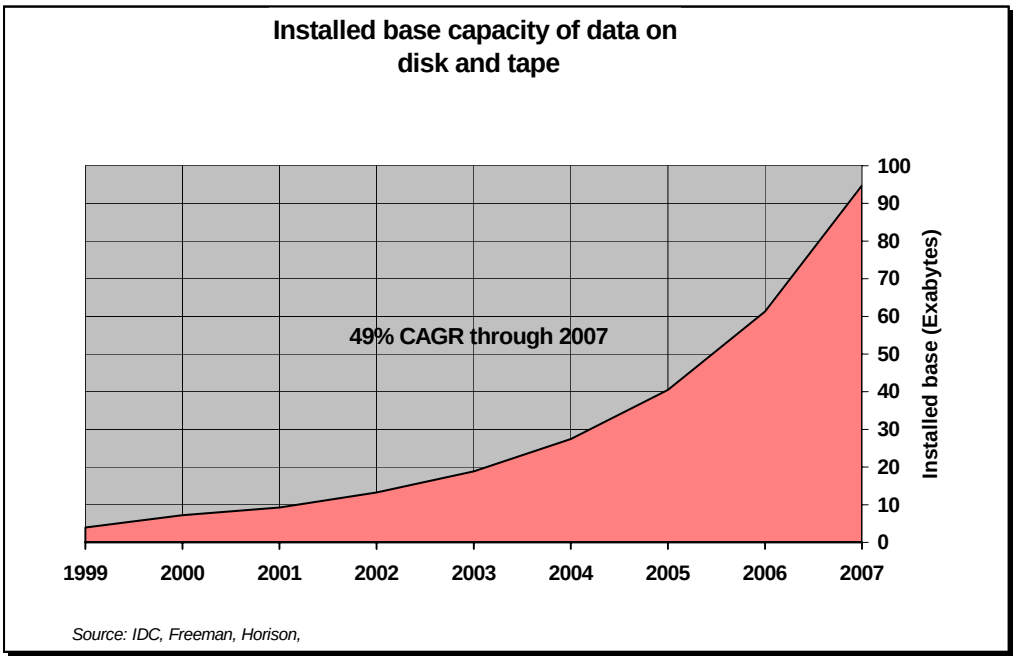
The Growing Importance of Archive

- ✓ **36% of all archived data reside on StorageTek solutions.**
- ✓ **The total installed base of all archived computer data is 15.28 Exabytes, or 80.3% of all stored data.** [1]
- ✓ **The projected compound annual growth rate of the installed base of all stored computer capacity is 49% through 2007.**

Why is Archiving So Important Now?

The confluence of an accelerating growth in the amount of data, regulatory compliance requirements, information lifecycle management, and a realization of the importance of business continuance is challenging the user to seek the most effective archive solutions. Given that the value of data changes over its lifetime, it is imperative that data are stored using the most cost-effective and reliable solution at each period of the lifecycle.

Data Growth—Several factors are driving the growth of digital data including growth in fixed content, email retention requirements, and the expanding sizes of all types of files such as email and databases. In addition, the lack of effective storage management policies results in the retention of old data with little or no value. The bottom line: more data are being accumulated and stored for longer periods of time.



1. Using IDC shipment data and cross-referencing a University of California at Berkeley study, it is estimated that 2.094 Exabytes of total installed base of computer data was stored on disk subsystems at year-end 2003. Horison Information Strategies reports four to fifteen times that amount is stored on tape media *(Storage Navigator 2004, p.11)*. Freeman Reports applied a factor of eight times in the study, resulting in 16.752 Exabytes of total data stored on tape, year-end 2003. One percent of "Online" storage, 30% of "Nearline" storage and 100% of "Offline" tape storage was allocated to archive, resulting in 15.275 Exabytes of "Archived" data, or 80.3% of the total.

Compliance requirements—mandatory compliance to a host of government regulations is forcing users to retain critical records for lengthy time periods. In most cases, severe penalties are imposed for noncompliance. A partial list of government regulations follows:

- **Health Insurance Portability and Accountability Act (HIPAA)—** Requires that all patient related information be retained and protected during patient's lifetime plus two years after patient's death.
- **Sarbanes-Oxley Act—**Requires the accounting industry to retain all audit correspondence of publicly traded companies for a minimum of four years after audit.
- **Securities and Exchange Commission Rule 17a-4—**Requires brokers and dealers to retain all customer financial statements for three years, trading account records for end of account life plus six years, emails and other correspondence for six years, and broker/dealer data for end of life of enterprise. Requires non-alterable media.
- **21 CFR Part 11—**Requires pharmaceutical, biotech, and lab companies to retain data related to manufacturing of drugs and pharmaceuticals for three years after distribution, and records related to manufacturing of biological products for 5 years after end of product life; and requires food processing companies to retain data for two years minimum after distribution.
- **OSHA—**Requires retention of medical information for 30 years of persons exposed to toxic substances.
- **Department of Defense 5015.2—** Relates to security and retention best practices resulting in certified records management solutions as applicable to Department of Defense operations.

Information Lifecycle Management—Archiving is a fundamental component of any Information Lifecycle Management (ILM) strategy. The archive is the final depository for virtually all data. A well understood principal of ILM is that the need to access data, and often the value of the data, declines rapidly as the data age. Thus space-efficient and cost-effective archive solutions are required to store the massive amounts of archive data.

Business Continuance—A heightened awareness of natural disasters, computer viruses, and terror attacks has focused attention on offsite storage of data to assure business continuance in the event the primary data is lost or destroyed.

Choosing the Right Archive Vendor

The selection of an archive vendor is as critical as the selection of an archive technology. Consider the following vendor selection criteria:

Breadth of technology offering—The ideal technology solutions provider offers a broad range of storage hardware and software technologies. Ownership of core technologies, such as storage recording and automation design, is key to controlling future developments and maintaining a leadership role. Appropriate strategic partnerships with leading third party partners should be used to augment the primary vendor's capabilities whenever required.

Ability to integrate Archive as part of an overall ILM strategy—ILM strategies and implementations vary widely from vendor to vendor. A vendor steeped in multiple hardware and software disciplines which are driven by a common set of in-house standards and directives, can integrate the archive function

into the ILM solution in a straight forward manner. Look also for a tape-centric vendor.

Industry position—Critical vendor capabilities such as in-house product development and design, world class manufacturing, global service capability, and availability of a full range of professional services are hallmarks of an industry leader.

Conclusion—Archive at an Inflection Point

What makes archiving so much more important now than ever before? Simply put, it is the **extraordinarily massive amount of data which must be archived and the quickening pace at which that amount of data is accumulating.** The challenge facing the storage industry is not only to provide reliable and cost-effective archive solutions today, but also to keep pace with rapidly escalating demands.

Market dynamics—Archive data accounts for more than 80% of all stored computer data. Business continuance concerns, government mandated long-term retention periods of critical data, and emerging corporate Information Lifecycle Management policies and practices have conspired to challenge traditional storage solutions. Best of class approaches and vendors must be employed to achieve workable archive solutions.

The opportunity—Archive data requirements are growing at a 50% per annum rate, with no slowing in that rate anticipated.

Tape is the ideal archive solution— Virtually all archive data reside on tape, with only a fraction of a percent on optical disk. Tape WORM capability is now available in most popular tape formats, providing permanent recording whenever required. For more than five decades tape has been the medium of choice for low-cost, long-term storage of archive data. Tape's aggressive technology development roadmaps and its continuing decline in cost per GB will assure that it will continue to play the lead role as the archive solution.