



ESG REPORT

Thought Leadership Series: Data Management Challenges in the Internet Computing Era

By Mary Johnston Turner
With Steve Duplessie

April, 2008

Table of Contents

Table of Contents..... i

Data Is Changing – and Growing 1

 Disruptive Changes in Data Creation and Usage are Driving Explosive Growth 1

 The File Storage and Management Situation Is Only Going to Get Worse 3

 Getting a Handle on File Management Requires New Processes and Tools..... 4

 SharePoint – the Double-edged Sword 5

 Holistic Infrastructure Virtualization Will Become Mandatory 5

 File Management Is Going to Be *the* Issue..... 6

The Bottom Line..... 7

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508)482-0188.

Data Is Changing – and Growing

Disruptive Changes in Data Creation and Usage are Driving Explosive Growth

Innovative technologies, global connectivity and new usage models have enabled an entirely new generation of computer users to create and share digital content faster and easier than was even conceivable just five years ago. Corporate computing environments, while lagging behind the consumer markets, are slowly but steadily moving into the realm of Web 2.0. Whether it involves an online community, social networking site, new media or collaboration, the commercial computing world will need to prepare itself for the rapidly evolving realities of business. Tools such as SharePoint, blogs, wikis, streaming media and other digital content creation and management applications are enabling organizations to redefine themselves in an almost real-time manner. New media content is being created for everything from training to marketing and becoming a mandatory component to everyday business. Whether it's in blog or video format, content is easier than ever to create—and data management will become harder than ever without significant changes in management tools.

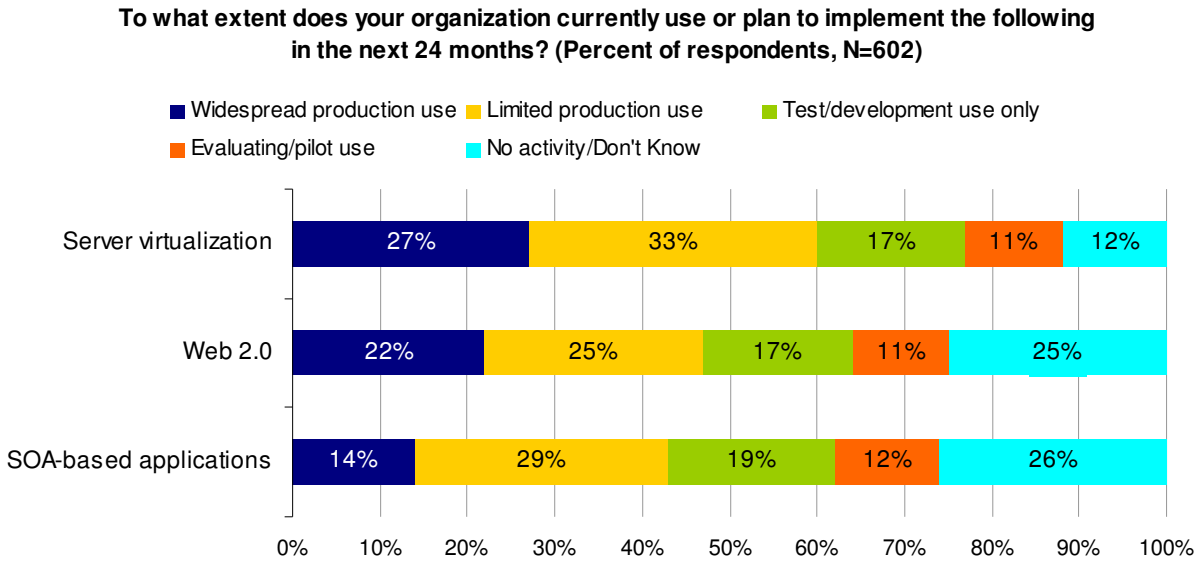
In this Internet Computing era, the very nature of data has changed. No longer primarily concerned with block-based, transactional data, new IT initiatives must deal with a new breed of data—one that is almost exclusively file-based. Furthermore, the files themselves are changing—becoming larger as the richness of their content increases. With the ability to create files becoming ever easier, it is no wonder that the amount of file data we must contend with is growing exponentially. Unfortunately, massive growth in the volume of file data and the already enormous complexity of enterprise infrastructures often lead to inefficient management and lost opportunities.

Basic data management has been a perplexing problem within IT organizations for many years. New business processes—coupled with the unrelenting growth of new, large file-based digital content—will undoubtedly necessitate a new approach. There is a clear need to re-evaluate processes predicated upon transaction-based computing of the past and focus on new ways to deal with the criteria of this new era. The way we handle file storage, backup, archive, search/retrieval, content delivery and collaboration will inevitably have to change—in terms of both infrastructure and management. In addition, infrastructure files in the form of virtual machine images, reusable web services, file systems and management databases are all proliferating as well. IT managers charged with protecting and managing the dynamic range of files face an almost untenable situation.

Recent ESG surveys¹ indicate that these often-disruptive new technologies are already operational in many corporate IT environments—and if they are not already operational, they soon will be. In fact, 27% of participants in a recent global survey stated that server virtualization is in widespread production use in their organizations, and another 33% indicated that it is in limited production use. Likewise, Web 2.0 applications such as SharePoint and other community-oriented content creation tools are in widespread production use in 22% of organizations polled and in limited production use in another 25%. Service-Oriented Architecture (SOA) is another significant factor, with SOA-based applications in widespread use by 14% of respondents and in limited production use among another 29% (Figure 1). Easier content creation capabilities and global connectivity combine to result in dramatically increased file-based content – requiring more and more storage capacity, causing even greater management concerns.

¹ ESG Research Report: *2008 IT Management Survey: Making the Case for Automation & Best Practices*, 2008

FIGURE 1. RESPONDENT PLANS FOR SERVER VIRTUALIZATION, WEB 2.0 AND SOA

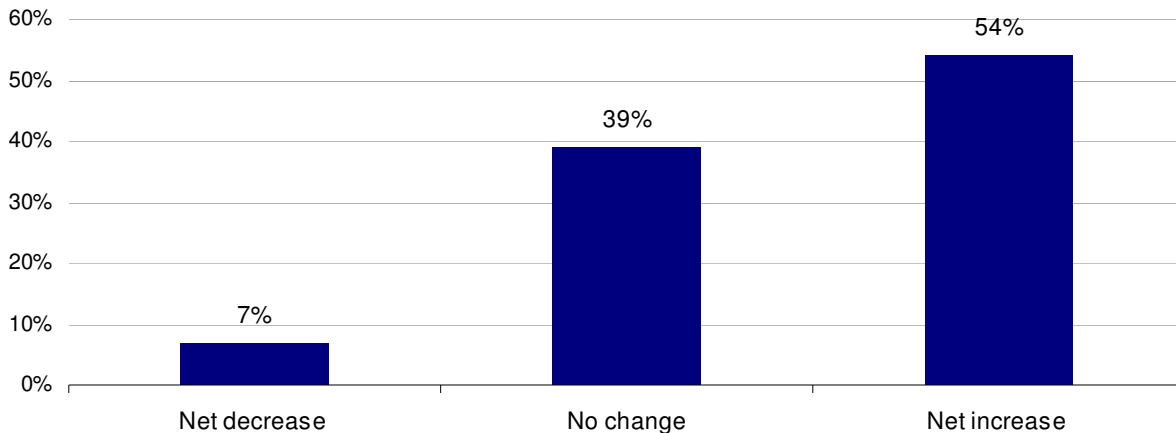


Source: Enterprise Strategy Group, 2008

Using server virtualization as an example, ESG’s analysis indicates that it significantly increases primary and secondary storage requirements. Over half (54%) of current users who were able to estimate their organization’s storage volume stated that their organization has experienced net growth in total storage volume after implementing server virtualization. Only 7% reported any type of net storage volume decrease (see Figure 2).²

FIGURE 2. IMPACT OF SERVER VIRTUALIZATION ON OVERALL VOLUME OF STORAGE CAPACITY

Since being implemented, what impact has server virtualization had on your organization’s overall volume of storage capacity? (Percent of respondents, N = 365)



Source: Enterprise Strategy Group, 2007

ESG believes that organizations reporting no change or a net decrease in capacity requirements might have benefited from storage consolidation or other infrastructure rationalization initiatives. It is also possible that these

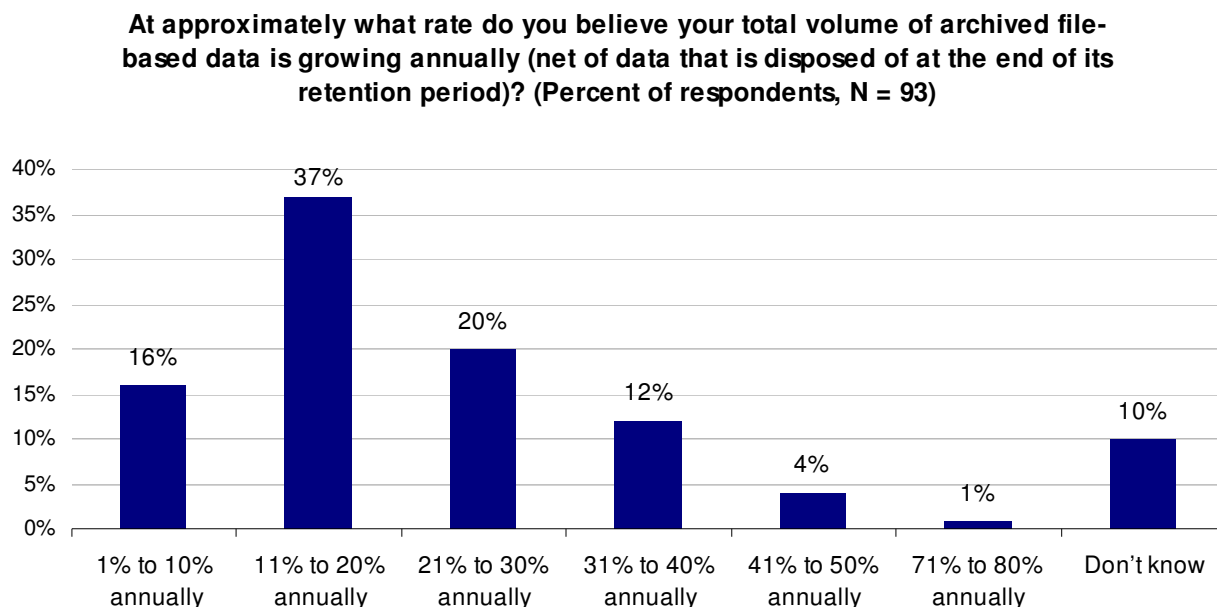
² ESG Research Report: *The Impact of Server Virtualization on Storage*, 2007

respondents might simply be in the early stages of server virtualization implementation and have not yet reached the tipping point where they see net demand for storage growing.

The File Storage and Management Situation Is Only Going to Get Worse

File-based content represents the greatest growth area of all data types in the new era. ESG's file archiving market forecast³ projects that total worldwide file archive capacity will increase from 7,119 petabytes (PB) in 2007 to 62,749 PB in 2012—a 55% compound annual growth rate (CAGR). These surveys also indicate that the growth in volume of archived file-based information exceeds all other categories—enterprises show approximately 10X growth in the volume of file data stored over the past two years, while SMBs report almost 13X in the same timeframe. Most customers anticipate that growth of file-based archives will continue unabated. Over the next several years, 37% expect the size of these types of archives to grow between 11% and 20% annually, while an additional 37% expect them to grow more than 20% annually (see Figure 3). Primary data may grow even faster as corporations begin to apply Web 2.0 and SOA technologies into their traditional IT environments.

FIGURE 3. ESTIMATED ANNUAL GROWTH RATE OF ARCHIVED FILE-BASED DATA



Source: Enterprise Strategy Group, 2007

As challenging as these rates might seem, ESG's analysis indicates that respondents actually underestimated the real growth rates of file data and related storage capacity requirements. For the most part, survey respondents providing these forecasts focused on traditional file types. They did not generally consider such file types as virtual machine or desktop images, reusable web services images or other infrastructure software images and file types. Furthermore, ESG's analysis shows that many customers significantly underestimate the size of files created by new digital content formats such as audio, video, image and streaming media.

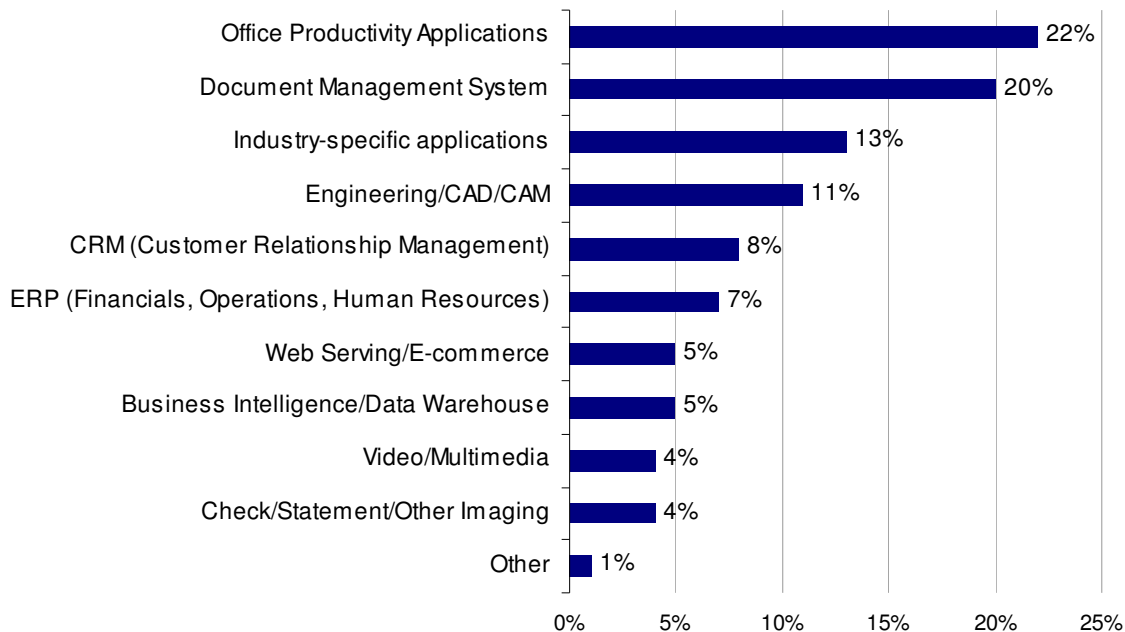
When asked what one single application is responsible for generating the most content, respondents cited office productivity applications (22%) and document management systems (20%). As might be expected, more structured applications such as ERP and CRM solutions generate less content because the core data supporting those systems is stored in structured database environments. Images used to document ERP and CRM

³ All subsequent references to ESG research are from the report: *2007 File Archiving Survey, 2007*

transactions do contribute somewhat to the amount of unstructured information (file data), but not as much as office productivity and document management solutions (see Figure 4).

FIGURE 4. APPLICATIONS GENERATING THE HIGHEST VOLUMES OF FILE-BASED ARCHIVING REQUIREMENTS

Of the following application types, which one would you say is creating the highest volume of archived file-based content? (Percent of respondents, N = 76)

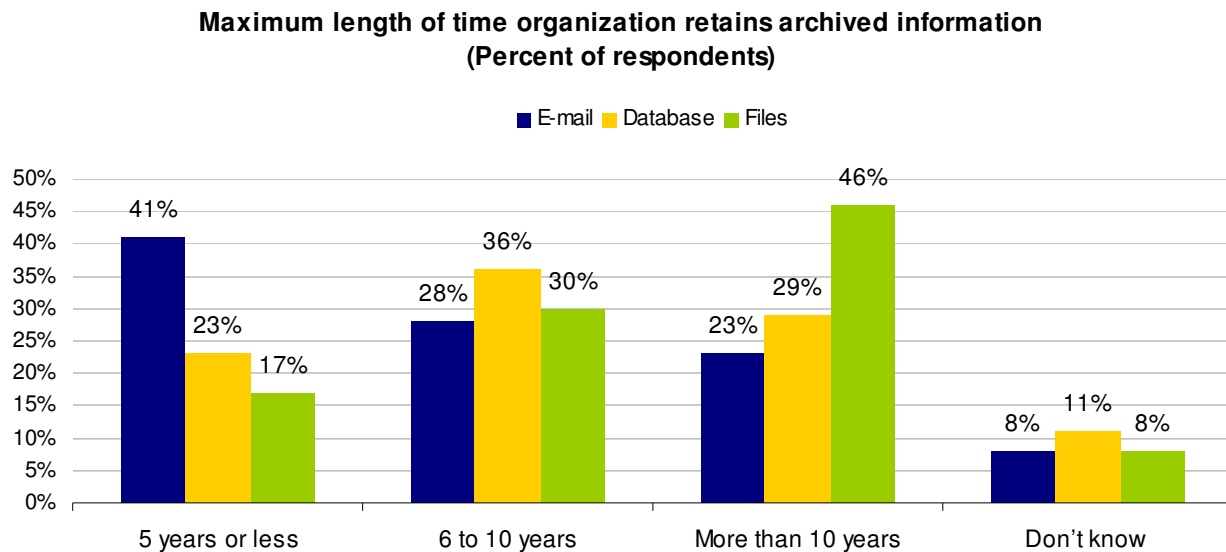


Source: Enterprise Strategy Group, 2007

Getting a Handle on File Management Requires New Processes and Tools

The survey indicates that data center managers are struggling to implement consistent, cost-effective file management processes, policies and tools. IT's ability to manage files has been much less effective than its ability to manage other applications such as e-mail and databases. Litigation support and e-discovery are two examples of current applications requiring existing data to be indexed and searched – which is relatively easy to do with structured and semi-structured data, but has proven daunting with unstructured file-based data. Applying policies to data for security or record retention can be even more difficult. Without consistent policies in place, it has been almost impossible to automate file management overall, let alone everyday functions such as file data migration, replication or backup. The existing processes are frequently tactical, manual, error-prone approaches that are time-consuming and create business risk. Worse yet, as file data growth rates continue to escalate, even well articulated, manual processes face the ever-increasing likelihood of failure or delay.

Another major concern is the fact that retention periods for archived files exceed all other content types. Forty-six percent of organizations say they must retain archived files for more than 10 years. In comparison, just 29% and 23% of organizations surveyed by ESG say they must retain archived database and e-mail information, respectively, for an equivalent length of time (see Figure 5).

FIGURE 5. MAXIMUM ARCHIVE RETENTION PERIOD, BY CONTENT TYPE

Source: Enterprise Strategy Group, 2007

SharePoint – the Double-edged Sword

IT managers especially have faced challenges in establishing clear policies for migrating file-based content to lower cost storage resources or for deleting it altogether. It seems as though end-users push for long-term, easy access to every version of every document, blog or wiki ever created. As file types proliferate and the sizes of files escalate, this situation will only worsen—and quickly. These high levels of archive data growth contribute directly to rapidly growing storage and administrative costs and, if information indexing and searching are unavailable, can increasingly hinder end-users from quickly finding the information they need.

SharePoint is quickly gaining corporate popularity as a universal virtual repository for file-based data. What is interesting is that while the presentation layer is a file management system, SharePoint's underlying intelligence resides on an SQL database. Today, SharePoint files and the database are unified, but future iterations will enable the database to contain only metadata. This means that SharePoint could end up being the file management system for files no matter where they exist—in SMB/CIFS or NFS shares anywhere in the enterprise. While that represents a possible boon for Microsoft and those suffering from ever-growing file management dilemmas, it also creates potential infrastructure issues.

Holistic Infrastructure Virtualization Will Become Mandatory

We have all witnessed the explosive growth of VMware and server virtualization, and while it is a critical part of the infrastructure, server virtualization alone cannot solve the problems ahead.

Server virtualization eliminates the physical ties binding an application, the operating system, and the associated server. Physical associations of that application to a specific machine have been at the top of the “stovepipe” infrastructure and are a significant reason why data centers are so complicated and difficult to manage. Any direct physical tie means any change state—whether a system migration or a failure—will disrupt the business unit requiring that application.

In contrast, server virtualization enables fluidity and mobility of the “virtual machine”—which consists of an OS image along with application image(s) and its data association or location. By mobilizing that application instance, administrators can create more fluid opportunities such that applications can continue to run without serious interruption, even in the event of failure or routine maintenance of the infrastructure. This VM fluidity theory works perfectly as long as all the necessary data associated with that VM is attached to every physical machine to which the VM can move. That requires the connectivity layer—or network—beneath the server layer

to connect everything above to everything below. If the data associated with an application happens to reside on a fibre channel disk array, the only way for that VM machine to retain access to the required data is to stay on a physical node connected to that array—which restricts mobility at the server layer.

Ethernet networks connect all systems in an organization. For applications requiring block data services, iSCSI has proven to be a viable option. File servers have garnered even greater popularity as virtual machine technology has become more widely adopted. By leveraging Ethernet, maintaining connectivity to data becomes significantly less restrictive. Regardless of the protocols or transport involved, the infrastructure you manage must be able to be as adaptable and virtual as possible.

Regardless of where virtualization occurs, it provides the same basic capabilities. Ensuring that all virtualization components work together seamlessly is critical to creating an intelligent, dynamic infrastructure. By default, virtualization technologies have two fundamental purposes:

1. Abstract the physical components beneath the virtual interface, allowing physical change states to occur without knowledge or interruption of the application above
2. Enable the automation of manual processes on the abstracted physical components

Going back to the server virtualization example, once the application is abstracted from the physical machine and operating system instance, it can move to other physical machines without normal restrictions that cause downtime. With storage virtualization, it might mean that the application reading and writing data from a virtual device or file system will not know (or care) that the data it is accessing resides on physical array B instead of the array it was on 10 minutes ago. As long as the connectivity/network can support dynamic change at either end of the wire, the application should never have to suffer the effects of movement or change.

File Management Is Going to Be *the* Issue

Eventually, infrastructure virtualization will be widespread and commonplace. This will allow us to eliminate some of the major problems in dealing with massive data growth, and provide the ability to scale out in any dimension dynamically. No matter how uncertain our needs might be, once capacity and performance issues are dealt with, we will face the true challenges of the Internet computing era—how we manage, protect and create value from our digital files.

To rein in the cost of file storage and improve end-user accessibility and compliance reporting, IT managers will need to define rules for policy-based, automated file movement. A practical example of this is implementing tiered storage for cost savings but pulling files back from a low-performance/cost tier to a high-performance/cost tier when necessary—as in the case of an e-discovery operation. IT managers will also need to automate for consistency and apply policies/rules for file placement, movement, encryption and retention. Performing these functions is difficult enough in a physical world—it will become even more difficult in a virtual world without improved intelligence and integration.

FIGURE 6. ONGOING FILE ARCHIVING CHALLENGES REPORTED BY CURRENT USERS

What ongoing challenges—if any— is your organization experiencing with respect to the management and archiving of file-based content? Please rank from 1 (greatest challenge) to 6 (smallest challenge)
1. Difficult to scale storage environment to meet file-based content growth
2. Poor search tools
3. Locating relevant content across the enterprise
4. Lack of clear organizational policies on data retention, etc.
5. Files are scattered across multiple devices (e.g., storage, servers, desktops)
6. Files are scattered across multiple geographic locations

Source: Enterprise Strategy Group, 2007

The Bottom Line

Getting your tactical act together is increasingly important, because file management will likely become even more difficult as time goes on. IT needs to consider how the world has changed in terms of creating and sharing digital content, and use the changes that are occurring as a reason to introduce innovative new processes to prepare for inevitable future developments by:

- a. Creating an environment that enables dynamic infrastructure fluidity
- b. Creating an environment that enables dynamic data fluidity, including file/volume migrations

Acknowledge the realities of your current environment, consider what is coming next, and begin formulating plans to get into a position to succeed. Leverage the momentum of corporate initiatives such as “Green IT” or server virtualization as a means to a larger end. Take advantage of consolidation efforts, ITIL initiatives, or any other corporate IT endeavor to re-evaluate your environment to put yourself in a better position to achieve effective file management.

The technologies might not all be perfect, but they do exist today and they do work. Although they might be limited to some degree, their horizons will expand rapidly, and they merit immediate consideration. You will be able to do everything that this paper suggests before the end of 2008.



20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com