White Paper

How to Make Oracle Databases Faster and More Efficient with Pure Storage

By Scott Sinclair, Storage Analyst and Nik Rouda, Senior Big Data Analyst

November 2015

This ESG White Paper was commissioned by Pure Storage and is distributed under license from ESG.

Contents

If Data Makes the World Go Round, Databases Are the Pivot Point 3

How to Approach Optimizing a Database Environment 4

Database Environment and Design Requires a Team, No Longer just the DBA 4

The Positive Impact of Solid-state Storage and the All-flash Array 4

Pure Storage and the Impact on Oracle Databases 5

Pure Storage and Impact of Flash Array//M 8

The Bigger Truth 9

If Data Makes the World Go Round, Databases Are the Pivot Point

Data has exploded. Today, much of the world is heavily instrumented, measured, and recorded, and ever so many events generate data. Some data is free form, coming spontaneously from human activities and loosely defined systems, while some is more tightly organized. Businesses are only as good as their ability to leverage data for insights into their operations, products, and customers, and competitive advantage will shift significantly to those that are most effective at harnessing data. Organizations of all sizes and across all industries will need to keep careful track of all transactions, recording information generated both internally and externally. Teams will also analyze this information both quickly for reactions in the moment, and again later more deeply and in broader context. Decision support for business managers is one compelling use case as it reduces executive guesswork and enables the organization to be truly "data-driven."

Today, the most common platform for managing enterprise data is the database, despite recent innovations such as Hadoop. Yet even within the realm of databases, there is much change with traditional, mature offerings—like Oracle's Database 12c, Microsoft's SQL Server, and IBM's DB2—now competing against newer alternatives like SAP HANA, MongoDB, Cassandra, and dozens of others. While the former category has had many years to build robustness and gain loyal followers, the latter group is forcing an evolutionary inflection point with new capabilities such as in-memory processing, distributed scale architectures, and DevOps-friendly versatility. Either way, the database is an essential centerpiece of many enterprise- and end-user-oriented applications, and the choices are about finding the database best suited to the specific need and the best IT architectures to support those decisions.

As the variety and functionality of databases changes, the diversity of the roles of the people engaging with them is also expanding. Application owners and developers are now joining the DBA because all are focused on building new offerings and capabilities around data. Focused on agile development of new business offerings, these groups are bringing new expectations and perspectives on the IT technology stack requirements, but they also may have less experience with database infrastructure and operations. As such, an open and ongoing conversation between DBAs, lines of business, and data center teams is critical to finding the optimal solution overall.

When designing almost any system or process, much of the architecture is determined by the bottleneck. This is no different when designing a data center environment, regardless of the workload. That being said, compensating for infrastructure bottlenecks can be even more important for highly transactional, performance-dependent, or mission-critical workloads. For over the past decade, the IT data center bottleneck was spinning disk. Now with solid-state storage, data center organizations are starting to realize a reality where that bottleneck has been removed, and are reaping the benefits.

Prior to the emergence of solid-state storage, IT organizations often went to great lengths to churn out the necessary performance from mechanical-based drives; deploying more spindles than necessary and even short stroking drives. As organizations replace the spinning media bottleneck with solid-state, the positive impacts can be felt throughout the data center. Over time, as the price of solid-state storage continues to decline and organizations gain a better understanding of the value solid-state provides, solid-state storage adoption will likely continue to increase. As this occurs, more IT organizations will continue to expand their solid-state storage footprints to more transactional workloads including database, analytics, and business intelligence applications.

Pure Storage, one of the leaders in the all-flash storage array industry, is working to extend the benefits of solid-state storage to a wider variety of workloads, including a broad range of analytics and database applications. With the modular design of Pure Storage's FlashArray//M array, combined with its Evergreen Storage program, Pure endeavors to further help reduce the costs of solid-state and ease integration of new technology. The resulting solution offers a compelling option for integrating the benefits of solid-state storage, and consequentially achieving more from your organization's database and analytics ecosystem.

How to Approach Optimizing a Database Environment

Database Environment and Design Requires a Team, No Longer just the DBA

Successful big data, BI, and analytics projects will inevitably require storage teams to be involved in planning, implementation, and ongoing administration. Overlooking the impact of storage on the overall performance, scale, and reliability of the stack will have negative consequences sooner or later. ESG research found 81% of survey respondents said engaging the storage team in planning is important or even crucial to achieving positive outcomes for these initiatives.[1] In fact, all IT disciplines are important, including teams specializing in servers, networking, security, application development, and more. These groups should discuss the mix of characteristics they need to provide for success, often including:

- Performance
- Scalability
- Availability
- Security
- Cost of ownership
- And more!

A very common consideration in choosing and managing database environments is the endless proliferation of data. More data from more sources leads directly to more and bigger databases, and this growth in number and size brings its own challenges for administration. A common approach to combat this trend is database consolidation. If the same data is copied in multiple environments, it could potentially be used instead from a single source—assuming that source can deliver enough performance at scale, and doesn't magnify risk too much (due to having more people dependent on one instance). Of course, a lot of data is also simply redundant or no longer relevant, so it can be reduced in the process of consolidation. This is a popular initiative, with ESG research finding 73% of respondents reporting that they are now pursuing consolidation broadly or narrowly—either way, it requires the right platform. While many organizations praise public cloud services for their seemingly infinite scalability on-demand, only 17% are considering cloud as their primary deployment methodology going forward. So most will need to rethink their on-premises configuration to accommodate both more data and hopefully fewer unique databases. As with much of IT, the foundational layer for a database is the storage…

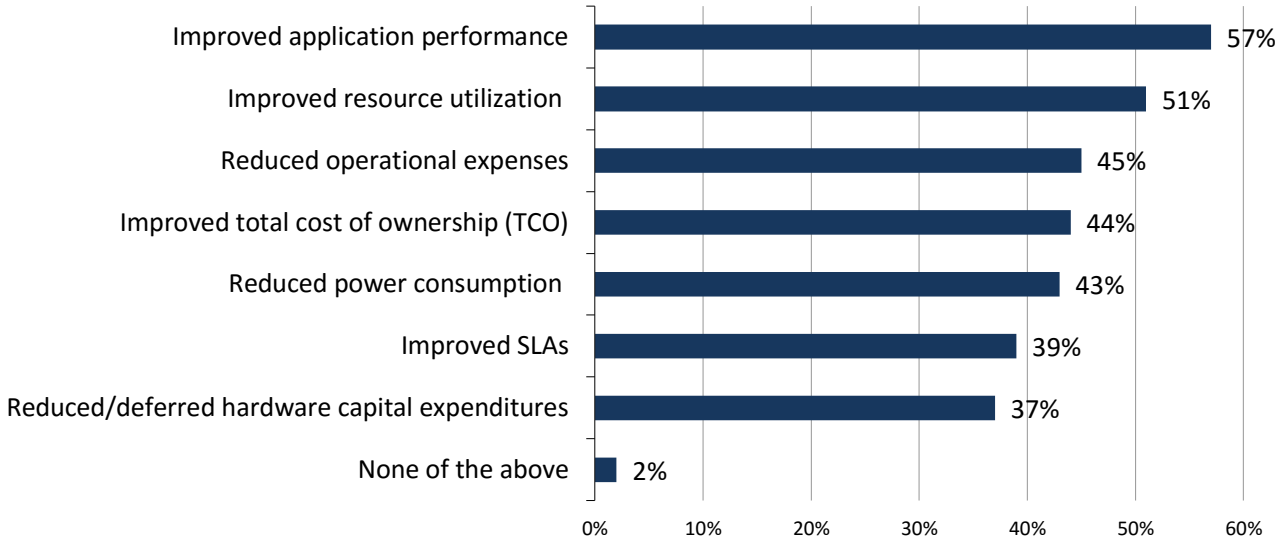
The Positive Impact of Solid-state Storage and the All-flash Array

In 2015, ESG conducted a research study investigating a variety of storage industry trends, including solid-state storage. This research study surveyed 373 IT decision makers responsible for their organizations' data storage environments. Current solid-state users identified in the study were asked to select the benefits their organizations realized as a result of deploying solid-state storage. While the most commonly identified response to this inquiry, improved application performance (57%), will not likely surprise anyone, the interesting responses were the ones that followed. The second, third, and fourth most-cited responses were improved resource utilization (51%), reduced operational expenses (45%), and improved total cost of ownership (TCO) (44%) (see Figure 1).[2]

[1] Source: ESG Research Report, *Enterprise Database Trends in a Big Data World*, July 2014. All ESG research references and charts in this white paper have been taken from this research report, unless otherwise noted.
[2] Source: ESG Research Report, *2015 Data Storage Market Trends*, October 2015.

Figure 1. Benefits Realized as a Result of Deploying Solid-state Storage

Which of the following benefits has your organization realized as the result of deploying solid-state storage technology? (Percent of respondents, N=181, multiple responses accepted)



Source: Enterprise Strategy Group, 2015.

These subsequent responses highlight the potential residual benefits that can be achieved by eliminating the data center bottleneck and deploying it with higher performing solid-state storage. In other words, thinking that adding solid-state to a storage environment offers simply higher performance is seeing only half the picture. Solid-state storage offers the potential for the remaining components of the ecosystem to operate more efficiently. Server and networking resources are no longer underutilized waiting for the spinning disk to respond. In turn, each database application or analytics workload can deliver more functionality. The net result can be, as indicated in the research, a net improvement to resource utilization and overall total cost of ownership. Those benefits can be critical in environments where significant budget outlays are required to deploy and maintain database licenses, which can be the case in Oracle database environments.

Pure Storage and the Impact on Oracle Databases

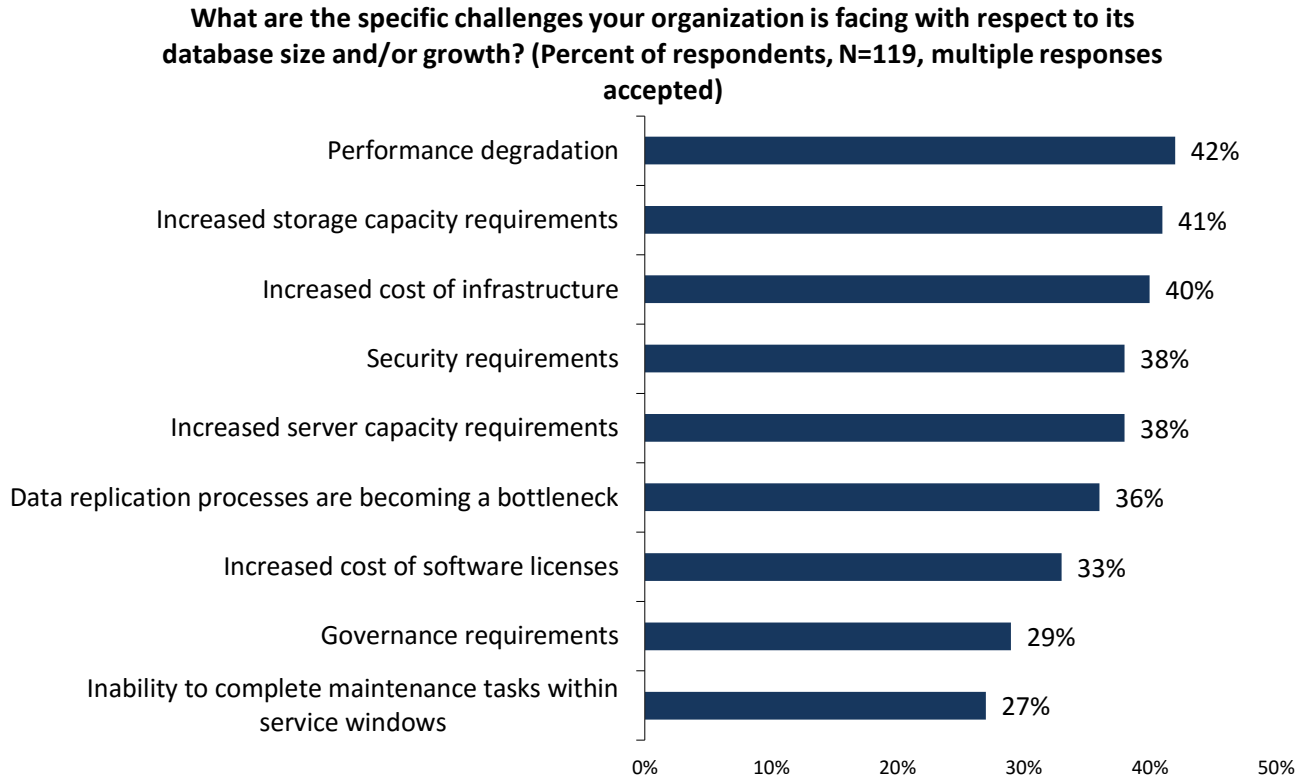
As already noted, database growth and proliferation is one of the most common scenarios faced, but what does this really mean? What are the impacts on a database that must be addressed? ESG research explored this topic and the results were reflective of the very same characteristics used for evaluation of the database environment (see Figure 2). Maintaining performance is the top concern as databases expand, with DBA-time-intensive techniques such as indexing only going so far in finding shortcuts to the specific data sought in an analytics operation. Oracle, for example, has added an in-memory option to accelerate access (borrowing a trick from some newer SQL and NoSQL databases)—in this case, effectively providing a columnar-oriented access route in server DRAM for analytics exercises alongside a typical row-oriented approach to recording transactions. Yet server memory is still comparatively more expensive than flash storage.

Other issues cited in the survey were that storage capacity and cost go hand in hand with the growth, which in turn drives the need for efficiency (unless you happen to enjoy infinite rack space and budget![3]). Again looking at Oracle, there are products that get quite larger indeed, with scale-up engineered systems offering staggering amounts of storage space, memory, or both. Yet even if these offer the resources and capacity sought, they come with a price that might be hard to swallow. A related problem is the high cost of application and database licenses, often priced

[3] This is unlikely.

per unit of resource (processors, database instances, etc.), and many traditional databases are not really well-priced for the extreme demands of the modern world and modern big data workloads.

Figure 2. Challenges Stemming from Database Growth and Proliferation



Source: Enterprise Strategy Group, 2015.

Despite these issues, ESG somewhat surprisingly found that 81% of IT decision makers say they are very satisfied or satisfied with their primary database vendor. So while many newer databases may be alluring for in-memory performance or scale-out economy on lower unit-priced commodity servers, these products are not going to displace a leader like Oracle overnight. More likely, they will find specific niches alongside the traditional RDBMS. Concerns about operational maturity and risk (including vendor viability) are still keeping these smaller vendors in the margins overall. This may change with time as disruptive start-ups gain ground, but it's unlikely that enterprises will move away entirely despite their challenges around topics like performance.

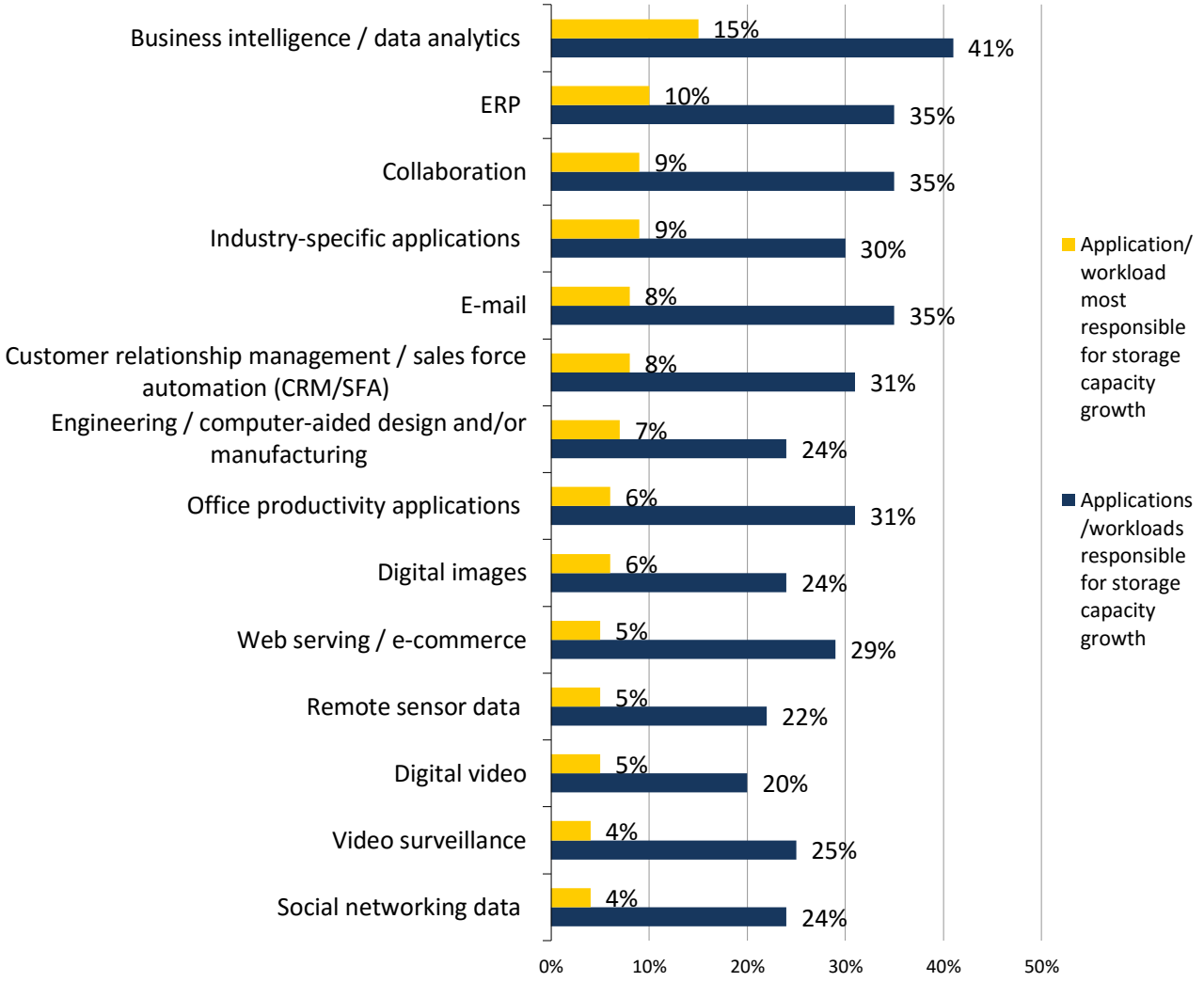
Instead, they will need to explore a better infrastructure to accommodate increasing demands and resolve the cited issues. Database indexing, tuning, and other local workarounds only go so far before simple constraints around physics and staff time dictate a new approach. Again, in-memory options for Oracle are very interesting, but are still inherently limited due to the significantly higher cost of server RAM at scale, as compared with flash and SSD storage today, never mind high-capacity spinning disks. With all of these pressures, the overwhelming interest in big data, BI, and analytics solutions is driving growth in database storage requirements overall, and in flash storage specifically for a balance of speed and cost.

Continuing with Oracle as an example, the Automatic Workload Repository (AWR) utility is a useful tool for understanding the hardware infrastructure limitations on the database. One concern might be whether the database is I/O bound. This utility can help look for the amount of time your database spends waiting for I/O operations to complete, as well as the most common types of I/O causing the wait. The more you can explore the specific bottlenecks, the better off you will be. Please don't be the DBA or developer hard-coding waits into your application to accommodate hardware constraints; there is a better way.

Just to drive home the point, consider ESG research into what will cause the most growth of storage in enterprise and mid-market companies over the next 24 months, shown in Figure 3.[4] Tada! The number one reported most common single cause and the most-cited most common of several factors turns out to be supporting business intelligence and data analytics. These are still most often built on databases. So the problem of managing scale and performance is going to get worse before it gets better. The thirst for more data will continue to drive the need for more effective and efficient storage.

Figure 3. Applications/Workloads Responsible for Data Growth over the next 24 months

Which of the following applications/workloads do you believe will be responsible for your organization's storage growth over the next 24 months? Which application/workload will be most responsible for storage growth? (Percent of respondents, N=373)



Source: Enterprise Strategy Group, 2015.

Although much of this discussion calls out Oracle databases, these issues are by no means only Oracle's problem. Most traditional databases will encounter the exact same demands and the exact same challenges. SAP HANA, Microsoft SQL Server, and IBM DB2 Blu Acceleration can all leverage their various implementations of in-memory operations, but with eventual limits on memory size and cost. Until next-generation scale-out databases can prove

[4] Source: ESG Research Report, *2015 Data Storage Market Trends*, October 2015.

themselves a bit more in mission-critical enterprise applications, most organizations will need to instead reconsider the hardware environment as part of the solution.

Pure Storage and Impact of Flash Array//M

For transactional database workloads, solid-state storage offers the potential for tremendous advantage. That being said, all solid-state storage offerings are not created equal. The delivery model of the solid-state storage plays a role in how workloads and IT organizations access its capabilities. For example, Pure Storage's Flash//M architecture offers a number of advantages specifically for database and analytical workload environments beyond some of the other solid-state alternatives.

- **Increased performance and reduced I/O latencies:** It may seem obvious, but reducing the latency enables a database workload to complete more transactions. The more transactions the workload can complete, the more effective that workload is and the greater the benefit will be to the business. For analytics and business intelligence workloads, this can significantly reduce the time required to analyze the data and generate reports. Business decisions can be made in a more timely fashion and reports fail less often. The performance headroom offered by all-flash array storage ensures enough IOPS to support any near-term application demand spikes and to support future demand growth expectations. The net result allows organizations to delay capital purchases for new server infrastructure. This can reduce the number of systems the data center has to manage, power, cool, and protect, which ultimately reduces costs and frees up budget for other activities.
- **Fewer application licenses and add-ons required:** With the storage bottleneck alleviated, server hardware and the application license can be freed to achieve their full potential. In many cases, this results in a reduction of server cores required, and consequentially the number of database licenses required to support the same business activity. Additionally, the automatic integration of data reduction techniques such as deduplication and encryption in the storage device removes the requirement to procure that feature as an add-on license for the application.
- **High resiliency and snapshots for development:** In addition to offering a highly available hardware design, along with other data resiliency capabilities such as replication, Pure Storage all-flash storage architecture offers the ability to quickly and non-disruptively create snapshots of active data volumes. For database environments, this ability can be critical to supporting active application development. In spinning media environments, concerns over impacts to performance often led database application development activities to leverage clones, rather than snapshots. Clones take longer to create and significantly increase the amount of capacity required. With the performance of all-flash storage, developers can leverage snapshots, greatly reducing the capacity requirement while speeding up accessibility for developers.
- **Future proof and performance scaling:** As storage technology evolves, storage architectures that can more easily integrate advances in processing, memory, and solid-state storage deliver greater value. Reducing the disruption of new technology can translate into significantly reduced management and operating costs. This is the idea behind Pure Storage's modular FlashArray//M architecture and its Evergreen Storage program. If a solution requires more performance, more memory, or more capacity, the system can simply be upgraded adding only what is needed. With Evergreen Storage, Pure Storage intends for IT organizations to deploy a FlashArray//M once and then upgrade the features and the hardware capabilities as needed without requiring a data migration. Pure Storage claims this modular capability allows the expected life span of each deployment to exceed ten years.

While this paper has referenced all-flash arrays as an example deployment model for solid-state, it is important to call out that other solid-state storage options can provide benefits as well. The impact can be dramatic for database and analytical workload environments regardless of the deployment model. That being said, trade-offs do exist and should be considered. For example, while hybrid arrays can offer access to solid-state storage at a lower entry price point, a large percentage of data will still reside on spinning media. For transactional workloads, predictable performance can be critical. All-flash arrays, such as those from Pure Storage, deliver predictability by only

leveraging solid-state. Additionally, the deduplication capabilities of Pure Storage's all-flash technology can significantly reduce the price per capacity allowing the predictable performance of all-flash storage to be delivered in a more cost-effective package.

The Bigger Truth

There isn't one single, well-defined path to success in life because many unknowns and unforeseen challenges lurk—an ideal database environment can be similarly confusing or fraught with peril. The good news here is technology vendors are continuing to innovate to find ever better approaches. The dynamic here is simple: More data will be housed in more and bigger databases, but performance cannot be compromised. If anything, the expectation for instantaneous, real-time analytics will grow as well. Finding the right storage platform that can survive and even thrive in these demanding conditions is imperative to business success. Pure is one of the leaders in solving the data challenges of the modern era.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel: 508.482.0188 Fax: 508.482.0218 | www.esg-global.com