BIG DATA, BIG DEMANDS

SCALE-OUT STORAGE MEETS ENTERPRISE REQUIREMENTS FOR BIG DATA

While the concept of big data is nothing new, the tools and technology are now in place for companies of all types and sizes to take full advantage. Enterprises in industries such as media, entertainment, and research and development have long been dealing with data in large volumes and unstructured formats—data that changes in near real time. However, extracting meaning from this data has been prohibitive, often requiring custom-built, expensive technology. Now, thanks to advancements in storage and analytics, all organizations can leverage big data to gain the insight needed to make their businesses more agile, innovative, and competitive.

There are a few important business drivers behind the growing interest in big data:

- **The desire to gain a better understanding of customers:** Companies want to learn which customers are most profitable, find ways to reduce customer churn, and discover ideas for improving profit-per-customer levels.
- **The quest for operational efficiency:** Companies want to gain a complete view into their business processes—be it quality control, manufacturing, or customer service—to improve efficiency and flow.
- **The need for risk management:** Specifically, companies want to improve IT security and reduce fraud, such as spotting advanced network threats and transaction anomalies.
- **The opportunity to innovate:** Companies want to discover new products, services, and business opportunities.

InfoWorld
Custom Solutions Group

COMPUTERWORLD
Custom Solutions Group

EMC²

"These are the same goals that companies have had for years, but with advances in storage and analytics, they can now extract the value that lies within all of their existing data quicker, easier, and more cost-effectively. They can even incorporate external data sources, enabling them to gain additional insights," says Generosa Litton, director of big data marketing with EMC.

To turn these business goals into realities, organizations must think about data management in new ways. Because big data is voluminous, unstructured, and ever-changing, approaches to dealing with it differ from techniques used with traditional data. Organizations that want to turn big data into opportunities should look for technology solutions that feature the following components:

- A versatile, scale-out storage infrastructure that is efficient and easy to manage and enables enterprises to focus on getting results from data quickly and easily
- A unified, agile analytics platform for structured and unstructured data with a productivity layer that enables collaboration
- Capabilities to become more predictive, driving action from insight and becoming more agile

With these essential components, enterprises can build infrastructures that deliver on the promises of big data.

"In our case, what big data means is creating value for individual stakeholders inside the company, who are working with large data sets and need to do critical analytics on that data," says Paul English, IT director of 3TIER, which performs analysis to help the global energy market manage renewable energy risk.[1] "We have built out a platform that supports our enterprise IT needs and high-performance computing needs to extract value out of massive data sets and deliver that value to customers. So we're extracting value and monetizing it."

> "In our case, what big data means is creating value for individual stakeholders inside the company, who are working with large data sets and need to do critical analytics on that data."
>
> — Paul English
> IT director of 3TIER

Big Data Transforms Renewable Energy

See how 3TIER uses EMC Isilon to put big data to work. With Isilon, 3TIER helps its customers balance risk with opportunity and capitalize on clean energy projects.
http://youtu.be/iqSNXZL8GFk

The Big Data Difference

Despite the many benefits it delivers, big data is putting big demands on organizations today, as it differs from traditional enterprise data in a few key ways:

- **It's voluminous:** Enterprises generate and collect large quantities of traditional data, but big data is often orders of magnitude more.
- **It's largely unstructured:** Big data includes Internet log files, scanned images, video surveillance clips, comments on a website, biometric information, and other types of digitized information. This data doesn't fit neatly into the rows and columns of a database, but accounts for roughly 80 percent of data growth today, industry experts say.
- **It's changing:** Big data often changes in real time or near real time— for example, customer comments on a website. This data must be collected over significant periods of time in order to spot patterns and trends.

What's more, enterprises are beginning to realize that to reap the full value of big data, they must be able to analyze and iterate on the entire range of available digital information. Mere snapshots of data do not necessarily tell the whole story or solve a particular business challenge. Yet, efficiently collecting and storing that data for iterative analysis has a significant impact on an enterprise's storage and IT management resources. Storage professionals need to find big data solutions that fit the bill, but don't strain already tight budgets or require significant investments in dedicated personnel.

[1] https://community.emc.com/community/events/live_events/theater_27

Due to these new big data demands, as well as the importance of handling information correctly, most enterprises consider managing data growth, provisioning storage, and performing fast, reliable, and iterative analytics to be top priorities, according to a recent study by market researcher Enterprise Strategy Group (ESG).[2] In this survey, conducted in March 2012, all 399 respondents said that big data is placing their current IT infrastructures under extreme stress; many are looking to build scalable infrastructures within their data centers. This stress dictates a new approach to storage.

"Now that enterprises realize they have this valuable data, a lot of them are looking into ways to leverage that data to help their businesses become more successful. They are also looking for ways to uncover nuggets of information that can identify new business opportunities, new ways of serving customers, and better methods for reducing costs," says Mike Noble, senior product marketing manager with EMC. "They also know they need to protect this data, safeguard it, and comply with regulatory requirements."

> "Now that enterprises realize they have this valuable data, a lot of them are looking into ways to leverage that data to help their businesses become more successful."
>
> — Mike Noble
> senior product marketing manager, EMC

Rethink Storage

Big data requires more capacity, scalability, and efficient accessibility without increasing resource demands. Traditionally, storage architectures were designed to scale up to accommodate growth in data. Scaling up means adding more capacity in the form of storage hardware and silos, but it doesn't address how additional data will affect performance.

"If you look at traditional storage architectures, RAID controller–based systems end up with large amounts of storage sprawl and create a siloed environment," says Eric Seidman, senior product marketing manager with EMC. "Instead, enterprises need to be able to achieve consolidation within a single, highly scalable storage infrastructure. They also need automated management, provisioning, and tiering functions to accommodate the rapid growth of big data."

EMC Isilon—The World's Fastest NAS System for Big Data

EMC Isilon Scale-out NAS has been certified by SPECsfs as the fastest NAS system in the world, based on validated test results of more than 1.1 million NFS IOPS and more than 1.6 million CIFS IOPS. Isilon's leading performance enables organizations to derive value from big data in a cost-efficient and timely manner. http://youtu.be/B1OT06ooGhc

Enterprises need storage architectures that are built with big data in mind and offer the following features:

- **Scalability** to accommodate very large and growing data stores, including the ability to easily add additional storage resources as needed
- **High performance** to keep response times and data ingest times low, thus keeping up with the required pace of the business
- **High efficiency** to reduce storage and related data center costs
- **Operational simplicity** to streamline the management of a massive data environment without additional IT staff
- **Enterprise data protection** to ensure high availability for business users and business continuance in the event of a disaster
- **Interoperability** to integrate very complex environments and to provide an agile infrastructure that supports a wide range of business applications and analytics platforms such as Hadoop

[2] ESG research report, July 2012: "The Convergence of Big Data Processing and Integrated Infrastructure"

The EMC Approach

EMC Isilon offers highly flexible, scale-out storage solutions ideally suited to the needs of big data storage and analytics. Powered by the innovative Isilon OneFS operating system, the Isilon scale-out network-attached storage (NAS) platform provides a highly efficient and reliable infrastructure that easily scales in both capacity and performance. It simplifies and reduces storage management to a few hours per week and optimizes resources with an automated, tiered storage strategy.

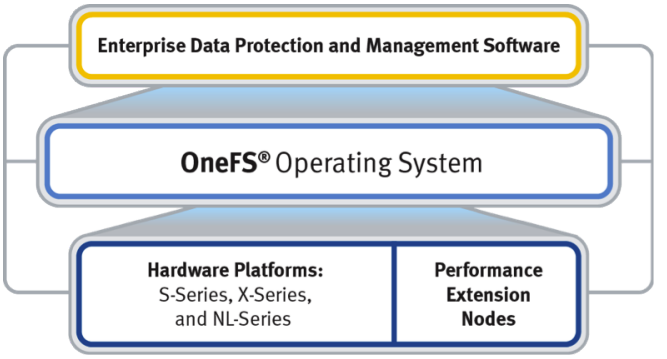
It also maximizes ROI with the highest storage utilization rates available on the market and reduces capital expenditures and operational costs.

Some key features of the **EMC Isilon scale-out storage platform** include the following:

- ✔ **Massive scalability:** Scales easily to more than 15 petabytes of storage within a simple-to-manage, single-file system and single volume
- ✔ **World-record performance:** Delivers more than 100 GB of throughput with a SPECsfs record-breaking 1.6 million IO/sec in a CIFS/SMB environment
- ✔ **Unmatched efficiency:** Increases efficiency with more than 80 percent storage utilization rates by preventing "hot spots" and automating storage management functions and tiering
- ✔ **Enterprise data protection and security:** Provides efficient data backup and protection, reliable disaster recovery, and WORM data retention
- ✔ **Management simplicity:** Scales capacity and performance without incurring an increase to OPEX, even within a single-file system, single volume, and global namespace
- ✔ **Operational flexibility:** Includes integrated support for a wide range of industry-standard protocols, such as NFS, SMB, HTTP, FTP, and iSCSI
- ✔ **Native Hadoop support:** Provides native Hadoop support by integrating the Hadoop Distributed File System (HDFS) with the EMC Isilon OneFS operating system for easy deployment of a comprehensive storage solution.

Conclusion

As the amount of unstructured data in the enterprise grows, companies are learning they need new approaches to managing that data. They require an efficient and scalable storage system to help them manage growth. In the ESG survey, the top three benefits cited for implementing scale-out storage were: 1) improved scalability, 2) improved performance (both I/O and throughput), and 3) improved data availability. Scale-out storage, paired with powerful analytics tools that can derive valuable insight from oceans of content, are the right combination for making the most of big data.



EMC Isilon scale-out storage solutions, powered by the OneFS operating system, provide users with a broad range of options to meet their big data storage needs.

Next Steps

To learn more about EMC's big data solution and how it can help you transform your business, please visit:

www.EMC.com/BigData

www.EMC.com/Isilon

twitter.com/emcbigdata

InfoWorld Custom Solutions Group

COMPUTERWORLD Custom Solutions Group

EMC²