



The Isilon essential guide series Oil & Gas Industry

Isilon® is the leader in clustered storage. Isilon's products are designed for data-intensive businesses and clustered computing environments, such as those found in oil and gas, media and entertainment, Internet services, manufacturing, life sciences, and government, which all produce, analyse or distribute large amounts of digital content, unstructured data and file-based information. Isilon helps world-class companies such as NBC Olympics, MySpace, Pratt & Whitney, Kodak EasyShare Gallery, GlobeXplorer, Technicolor, Movielink, and E! Networks meet the demands of their rapidly growing storage environments and accelerate their workflows.

Isilon Systems EMEA
59-60 Thames Street
Windsor, Berkshire
SL4 1TX, United Kingdom

t: +44 (0) 1753 272290
f: +44 (0) 1753 272291
www.isilon.com



Data storage requirements in the Oil and Gas industry are rising rapidly!

Oil and Gas is one of the most research intensive and technologically advanced industries in the world. Although many companies have only recently discovered the power of clustered computing, the Oil and Gas industry has harnessed distributed computing for several decades to enhance data analysis and identify new resources.

Advancements in underground imaging and visualization technologies have allowed exploration companies to collect an unprecedented amount of information to determine the best location for cost effective physical exploration. Although the increased amount of information collected has allowed companies to target discovery efforts with extraordinary accuracy, the sheer magnitude of data present a number of challenges.

Today, to get an accurate picture of what lies below the surface, an average seismic survey requires 10's to 100's of terabytes of information to be generated. With exploration activities rising to meet increasing worldwide demand, the Oil and Gas industry requires a combination of high end processing power coupled with scaleable storage systems.

With extreme competition within the industry, many organisations are developing architecture that will provide a competitive edge and more importantly, satisfy the potential terabyte capacities that will be necessary to meet the exploration challenges of the future.



Limitation on existing technologies.

Today's oil and gas visualisations are building in 3-D and 4-D resolutions and are often two to three times the resolution of previous generations of seismic data volumes. To enable its users to accurately interpret these massive data sets, Oil and Gas companies will build modified versions of these huge data sets to further analyse using specialist applications. These multi terabyte data sets are straining traditional storage systems, which were designed primarily for structured data with small file sizes and high transaction rates such as relational databases and transactional applications.

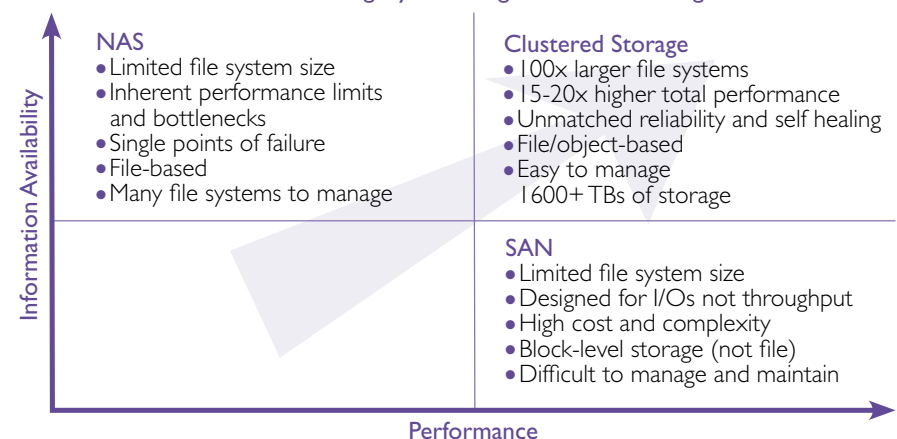
These data sets have unique characteristics for which traditional storage systems were not designed,

including large file sizes and data volumes, high throughput requirements, read intensive access patterns and high concurrent file access.

For lack of better alternatives, many companies have attempted to meet their need to manipulate large data files by extending traditional storage systems designed for structured transactional or text-based data.

Even the newest NAS and SAN systems employ architectures with inherent limitations that lead to enormous management complexity due to "islands of storage", limited scalability, performance bottlenecks, poor availability and high costs that are created when such systems are used for large data sets.

A New Category of Storage: Clustered Storage



Introducing clustered storage – Why is it the fastest growing storage technology?

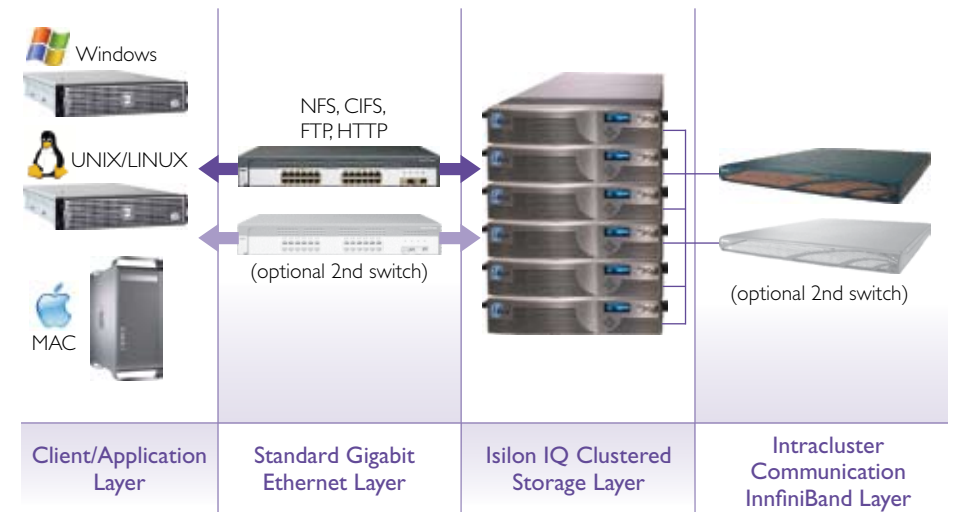
For any organisation that deals with seismic data and simulation, a new type of storage infrastructure is required that is designed with the content in mind. Clustered storage is the fastest growing type of new storage technology for the oil and gas industry because it was designed from the ground up to fit the needs of users with large data files.

The concept of clustered storage mirrors the shift from single big mainframes of the 1990's to computing power delivered from racks of low cost application servers that is popular today. Applications servers are connected together by network switches and load balancer to distribute the workload between different servers based on utilisations and availability. In this methodology, if a company needs more of a certain type of application, it is relatively easy to just add another low cost server to the pool and capacity is increased. Additionally, if a server should fail, then the availability of the application can continue due to the redundancy of the architecture.

Clustered storage works on a similar principle although the latest iterations place intelligence in each storage appliance which allow the cluster to self manage features such as capacity balancing, maintaining optimal transfer rates and error correction.

Unlike NAS, clustered storage is intelligent and provides a common single name space that can be treated just like a locally attached disk. When a new storage appliance (node) is added the cluster, the cluster itself redistributes the data to share it across all the other parts of the cluster and makes a number of copies of the file so that losing a node due to disk or chassis failure has no impact on the integrity of the data and allows the cluster to keep functioning.

Isilon IQ Network Architecture



Comparing the costs – Clustered Storage comes out ahead!

In the broadest of scales, as the amount of data under storage grows – clustered storage becomes the most cost effective option. For capacities under 1 Terabyte, direct and network attached storage is probably a good choice. Once you get over 4 terabytes, then Clustered and SAN is a better option.

At 50 terabyte plus or when you need to access large files of 500mb or greater, clustered storage offers unmatched cost versus capacity, especially in respect to the number of staff required to manage the storage pool.

At a more tactical level, some key financial considerations should be made clear:

Direct Attached and Network attached storage is initially the cheapest to install. Often used as a “quick fix” when companies need to deploy storage to fulfil an immediate need such as a new application or migration. However, both technologies have significantly higher costs in management time – especially when a drive fails and needs to be rebuild. The long-term cost of both technologies is also the highest as the data capacities grow past single terabyte figure with the formation of standalone islands of storage each requiring individual management resource.

SAN is true storage architecture with improved performance and ability to scale. One of the largest SAN's implementation at NASA is around 120 terabytes but the downside is a much higher initial and ongoing management costs, as the technology is not based on Ethernet but instead requires fibre channel. Additionally, SAN's are essentially dumb networks with the intelligence supplied by third party add-on software that is typically priced to match a perceived higher cost of the overall solution. The last cost issue is the ability to scale storage separately from performance. With SAN, the two are intrinsically linked and boosting just performance or capacity is not cost effective.

Clustered storage takes the simplicity of NAS and combines it with SAN like structure with a number of technology breakthroughs that reduce both initial capital cost and more importantly – ongoing management costs. In terms of capacity, the largest deployed Clustered Storage installations within commercial organisations like Kodak have exceeded the Petabyte barrier (1000 terabytes) and eclipse even the NASA SAN implementations.

By using either Ethernet and higher than fibre performing InfiniBand, the initial capital expenditure for a 15+ terabyte clustered storage solution is on average 40% less expensive than SAN. At 100 terabytes, the savings jump to over 50%.

Leading edge Clustered Storage vendors such as Isilon have all the management, error correction and balancing software built into the unit requiring no additional third party software expenditure and both storage or performance can be independently scaled.

Feedback from Clustered storage customers confirm that management requirements are considerable less than both NAS and SAN due to the self healing, balancing and error correcting features that are designed into Clustered storage and not bolted on via additional software.



How does clustered storage work?

Isilon IQ clustered storage is based on a unique modular design that combines a powerful server and up to 9 terabytes of disk capacity in each dense, self-contained storage node with up to XX nodes working together in a single cluster. This modular design combines intelligent software and smart hardware in a fault tolerant chassis.

By intelligently distributing content across the nodes in a cluster, Isilon IQ creates a shared pool of storage that can be viewed, managed and accessed from a single point. The smart software in Isilon IQ automates many previously manual tasks and provides flexible and incremental scaling, eliminating barriers for new projects.

At the heart of Isilon's clustered storage solution is Isilon's OneFS™ patented distributed file system. It combines the three layers of traditional storage architectures; file system, volume manager and RAID into one unified software layer, creating a single fully symmetrical file system that spans all nodes within a cluster.

OneFS uniquely stripes files and meta data across multiple storage nodes within a cluster, an improvement over the traditional method of striping content across individual disks within a single storage device or volume.

OneFS provides each node with knowledge of the entire file system layout and where each file and parts of files reside. Accessing any independent node gives a user access to all-content in one unified namespace, meaning that there are no volumes or shares, no inflexible volume size limits, no downtime for reconfiguration or expansion of storage and no multiple network drives to manage.

Performance:

Isilon's clustered architecture and OneFS distributed file system easily scales from 4 to more than 1600 terabytes of capacity running at up to 10 gigabytes/second of throughput all within a single file system. With Isilon's built-in TrueScale™ technology, storage capacity, system performance and throughput can be scaled linearly or independently to meet the most demanding capacity and performance requirements.

Configuration:

The top of the range IQ 9000 uses standard Gigabit Ethernet for front-end connections to Windows, UNIX or Macintosh servers and clients via Industry standard file-level network protocols including NFS, CIFS, FTP, HTTP, SNMP and NDMP allowing Isilon IQ to easily interoperate with existing systems. Intracluster communication uses high-performance, low-latency InfiniBand® via any third party InfiniBand switch.

Resilience:

Each node in an Isilon IQ cluster is a peer and can handle a request from any application server to provide the content requested. If any one node were to go down, any other node could fill in, thereby eliminating single point of failure.

Isilon IQ implementations can withstand the loss of multiple disks or entire nodes without losing access to any content. OneFS's unique FlexProtect-AP feature utilises Reed Solomon "n+1 and n+2" error correction codes, parity striping and mirrored file striping from 2 to 8 times that spans multiple nodes within a cluster. These policies can be set at any level, including cluster, directory, sub-directory, or even at the individual file level. Additionally, these policies can be changed at any time from a simple WebUI — even while the system is in production and fully available.



With Isilon, all files are striped across multiple nodes within a cluster; no single node stores 100 percent of any file, and if a node fails, all other nodes in the cluster can still deliver 100 percent of the files without interruption. Specifically, “n+2” double ECC error correction allows for multiple failures of disks or even nodes within a single cluster and file system. Isilon IQ is the only clustered storage solution to offer this level of data protection across a single file system in a clustered architecture.

In the event of a failure, OneFS automatically re-builds files across all of the existing distributed free space in the cluster in parallel using the backend InfiniBand connectivity, eliminating the need to have the dedicated “parity drives” typically required with most traditional storage architectures. OneFS takes advantage of the cluster by leveraging all available free space across all nodes in the cluster to rebuild data. By utilising this free space while also drawing on the multiple processors and compute power of the cluster, data can be rebuilt five to ten times faster when compared to traditional architectures.

How are organisations in Oil and Gas benefiting from Clustered Storage?

A growing number of companies within the oil and gas industry such as Brigham Exploration, Perenco and Penn Virginia have already deployed Isilon storage in-house. Many more are using Isilon powered solutions offered by service providers such as Arcis, Kelman Technologies, Seismic Exchange and Tricon Geophysics.



Perenco

The Perenco Group began operations in the oil and gas industry in 1975 as a marine services company based in Singapore. Anticipating the 1986 oil crisis, Perenco then expanded into the upstream business, acquiring several proven onshore oil and gas fields in the United States, applying secondary-recovery techniques to enhance production. After divesting its drilling interests in 1992, Perenco succeeded in building a worldwide portfolio of assets through a series of production acquisitions. Perenco's growth is determined by combining acquisitions of producing fields and undeveloped assets as well as high-value exploration.

Within the Perenco Geosciences Department, engineers and scientists from a variety of disciplines and International backgrounds exchange ideas, share expertise and customise

their skills, enabling quick assessment of each new project. The information technology aspect of the Geosciences Department is based in the Group's London headquarters. Data collected from projects all around the world is analysed by experts using high performance workstation connected via gigabit Ethernet to a number of standalone NAS server. Although offering reasonable performance, the routine management of the storage infrastructure and rebuild times after a drive failure was a constant burden for the small IT team.

As Jonathon Carey, IT Manager explains, "The old systems was cumbersome. Moving data around and managing LUN's was time consuming and after a failure, drive rebuilds would routinely take a couple of days."

"Isilon has delivered exactly what it promised by substantially reducing our management overheads, completely eliminated manual drive rebuilds while giving us a platform to easily and quickly expand capacity."

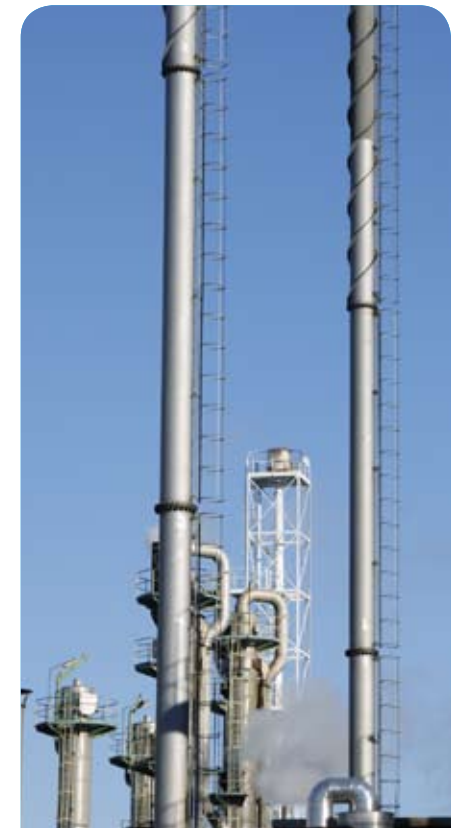
Jonathon Carey, IT Manager for The Perenco Group

So with the NAS boxes due for replacement as part of planned upgrade cycle, Carey and his team decided to explore alternatives that would allow them to both reduce management overhead and build a better storage architecture to match the 50% year on year growth in data that Perenco had experienced over the last few years. Following extensive market analyses, Perenco selected Isilon IQ as a replacement for its legacy storage systems.

The IQ cluster exceeded the initial requirement of Perenco as it natively supports all its Linux, Solaris and Windows workstations without the need for local clients. In addition, the ability to independently scale both performance and capacity would provide the Geosciences team with more flexibility if they needed to deploy any new number crunching applications.

The new cluster went live within a few weeks of the initial demonstration and has delivered 100% uptime to the 50 workstations within the Geosciences department. "Overall, the Isilon solution has been pretty uneventful which is a good thing," comments Carey, "Storage should be simple and Isilon has delivered

exactly what it promised by substantially reducing our management overheads, completely eliminated manual drive rebuilds while giving us a platform to easily and quickly expand capacity."



Kelman Technologies Inc

Kelman Technologies Inc. (KTI) is a publicly traded Canadian company with offices in Calgary and Toronto in Canada, Denver, Houston and Oklahoma City in the United States, and London and Tripoli. KTI provides a full suite of seismic processing and online data management and archival services that its customers use to locate oil and gas reserves.

KTI is a specialist in the analyses of seismic data, collected in marine and terrestrial environments using geophones (advanced microphones) to record vibration information from strategically placed explosives. This raw data, which can exceed 10 Terabytes for each sampled area, is collected on tape systems and then sent to KTI for processing.

After installing a state-of-the-art Intel Linux high-performance clustered computing farm, KTI discovered that its legacy storage systems, which consisted mainly of low-end RAID sub-systems, were inadequate to keep up with the speed of its growing computing systems. As Pat McKenny, V.P. Geophysical Processing, Kelman Technologies explains, "We rigorously evaluated a number of storage solutions and ultimately chose Isilon because it offered the most compelling combination of high performance and low capital acquisition costs. The performance and manageability of the Isilon IQ storage system has provided Kelman with a huge business advantage."

Kelman Technologies initially deployed more than 25 Terabytes of Isilon IQ 1920i clustered storage at its Houston, Texas location to meet the massive data throughput requirements of its high-performance computing clusters. Isilon IQ clustered storage enabled the company to achieve extremely high concurrent data throughput while removing the burdensome management constraints of traditional SAN and NAS storage systems.

In addition, because of the superior speed of the Isilon clustered storage solution, Kelman was able to implement its cutting-edge depth migration technologies and win previously unattainable contracts with large international oil and gas companies without expanding its IT staff.

"Isilon offered the most compelling combination of high performance and low capital acquisition costs."

Pat McKenny, V.P. Geophysical Processing, Kelman Technologies

For more information about Isilon products and services visit www.isilon.com or call +44 (0)1753 272 290 to request a copy of the Isilon White Papers 'The clustered storage revolution' and 'Uncompromising reliability through clustered storage'.

