Technical Report

IHS Kingdom on NetApp FAS Storage Systems

Application I/O Characterization and Sizing

Osama Qazi, NetApp
Ken Moore, IHS

May 2016 | TR-4510

Abstract

In this document we aim to provide an understanding of the upstream oil and gas interpretation application IHS Kingdom and describe its I/O characteristics. We share our findings based on the joint testing done with IHS and translate these for storage administrators in successfully architecting, deploying, maintaining, and growing their NetApp® storage environment.

Performance and scalability with always-on availability are top of mind due to data proliferation and increasing scale and complexity of interpretation and seismic data. Therefore, we comprehensively discuss how a storage solution based on clustered Data ONTAP® meets and exceeds these requirements. Last, we include sizing guidelines for architecting the appropriate storage system for a given Kingdom environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Seismic interpretation is the extraction of subsurface geologic information from seismic data. Seismic data, represented by seismic wavelets, is like an ultrasound that is used to record images of the rock layers below ground. These recordings carry the geologic information inside.

It is the job of the seismic interpreter or geoscientist to distinguish various elements in the geologic information recorded in the seismic data. These elements may include continuity of reflections indicating geologic structures, variability of reflections indicating fluids and reservoirs, and noise of various kinds.

In addition to seismic data, there are other types of field data and GIS data that are leveraged by the geologists, geophysicists, and engineers to create reservoir models in an effort to gain better understanding of the subsurface that eventually leads to informed well placement, reserves estimation, and production planning.

The IHS Kingdom application contains a suite of toolsets to facilitate the engineers' work from looking at the seismic recordings, inserting the well logs information, and manipulating seismic attributes to characterizing the subsurface and optimizing the extraction of hydrocarbons.

The NetApp clustered Data ONTAP 8.3.1 software provides the necessary data management, efficiency, availability, and performance to support interpretation software and workflows while masking the underlying complexity of heterogeneous data sources from the field, many different data types, and collaboration between hundreds of users and departments within the organization.

In this document we take a deeper look at the I/O characteristics and requirements of the Kingdom application and determine the best practices and sizing guidelines for deploying an efficient and well-performing NetApp storage system.

2 Overview of IHS Kingdom

IHS Kingdom is a powerful seismic interpretation platform and is widely deployed throughout the world in the oil and gas exploration and extraction industry. Kingdom maximizes the efficiency and productivity of seismic interpreters. It provides an integrated geoscience environment tailored for large and complex geoscientific workflows. The focus is on increasing productivity, and therefore the performance of the application is of utmost importance. We try and learn the I/O profiles of the various workflows and determine the best practices for deploying Kingdom on a NetApp storage system to provide the best performance.

2.1 Application Deployment

IHS Kingdom 2015 is a workstation-based application that runs in a Microsoft Windows 7 operating system. Many of the workflows in the application also require rendering seismic volumes in 3D; therefore, these workstations need to be equipped with supported NVIDIA GPUs. Moreover, the seismic interpreters and geologists typically require two monitors for viewing the 3D subsurface representations and the 2D/3D seismic volumes.

The configuration of the workstation and sizing depend on the size of the Kingdom project on which the users are working. Table 1 shows some guidelines from the Kingdom 2015 release notes.

Table 1) Workstation hardware requirements.

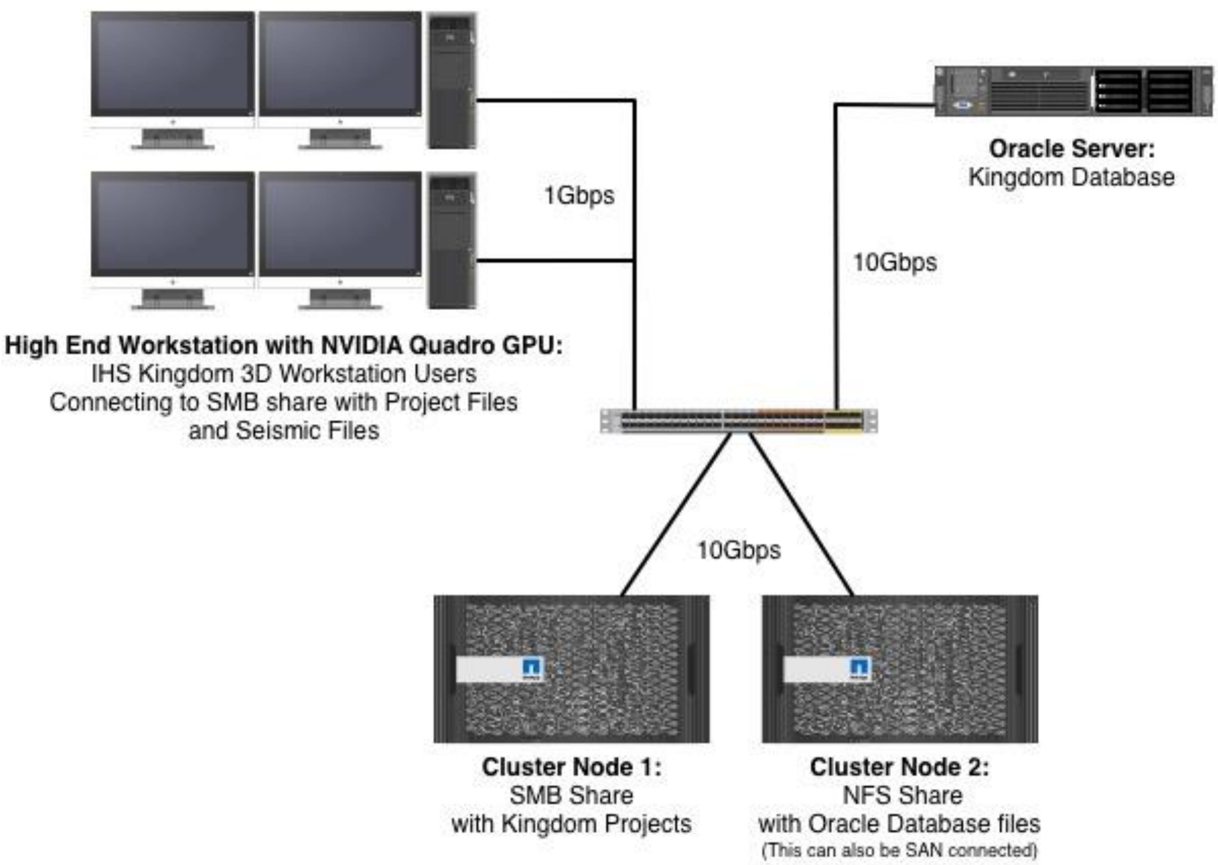
Project Size	Description	Minimum Requirement
Large regional (~50GB)	• More than 200,000 wells • 100–1,000 x 2D lines • >10,000 sq. miles of 3D	32GB RAM 16 CPU cores

	seismic data	
Large (~10GB)	• 50,000–200,000 wells • >500 x 2D lines • 200–10,000 sq. miles of 3D seismic data	16GB RAM 8 CPU cores
Small (~5GB)	• <5,000 wells • <500 2D lines • 200 sq. miles of 3D seismic data	8GB RAM 8 CPU cores

Kingdom 5 requires the use of a SQL Server Express for approximately four concurrent users, SQL Server, or an Oracle database for larger deployments.

Figure 1 shows a typical deployment scenario.

Figure 1) Typical customer deployment.



2.2 Application Project Data Layout

Kingdom project data can be broken down in two components: unstructured and structured.

Figure 2) A typical Kingdom project folder.

Name	Date modified	Type	Size
Annotation_V86	12/30/2010 1:29 PM	File folder	
DynamicDepthConversion	12/14/2015 5:18 PM	File folder	
EdmxFiles	12/14/2015 5:19 PM	File folder	
ErrorLogs	12/14/2015 8:56 AM	File folder	
Log	12/14/2015 8:56 AM	File folder	
My Collections	12/14/2015 5:19 PM	File folder	
Old_Project_Files	12/14/2015 5:19 PM	File folder	
PAKnotesTemplates	12/14/2015 8:56 AM	File folder	
Performance_Scripts_74	1/14/2016 4:26 PM	File folder	
PreferredLists	12/14/2015 8:56 AM	File folder	
Public	12/14/2015 5:10 PM	File folder	
Scripts_74_Version2	12/14/2015 8:56 AM	File folder	
Scripts_75	12/14/2015 8:56 AM	File folder	
SEGY	3/7/2016 11:13 AM	File folder	
TKSWorkTree	12/14/2015 8:56 AM	File folder	
user2	1/15/2016 4:02 PM	File folder	
user3	1/15/2016 4:03 PM	File folder	
user4	1/15/2016 4:05 PM	File folder	
user5	1/15/2016 4:04 PM	File folder	
user6	1/15/2016 4:04 PM	File folder	
seismic.1	2/29/2016 1:08 PM	1 File	38,177,123 KB
Seismic_.1	2/19/2016 3:48 PM	1 File	9,432,581 KB
31GBsegy.sgy	2/19/2016 1:18 PM	Viewer File Type	115,092,905 KB
Author.config	2/19/2016 3:34 PM	CONFIG File	2 KB
AuthoredMistieAuthorControlData.dat	2/8/2013 2:32 PM	DAT File	1 KB
AuthoredMistieControlData.dat	12/30/2010 1:29 PM	DAT File	1 KB
AuthoredMistieData.dat	12/30/2010 1:29 PM	DAT File	1 KB
AuthorReport.html	6/30/2004 5:02 PM	Firefox HTML Docu...	57 KB
Copy of SMT_LoadTest70.mdb_050421	4/18/2005 11:42 AM	MDB_050421 File	1,198 KB

2.2.1 Unstructured or File-Based Project Data Components

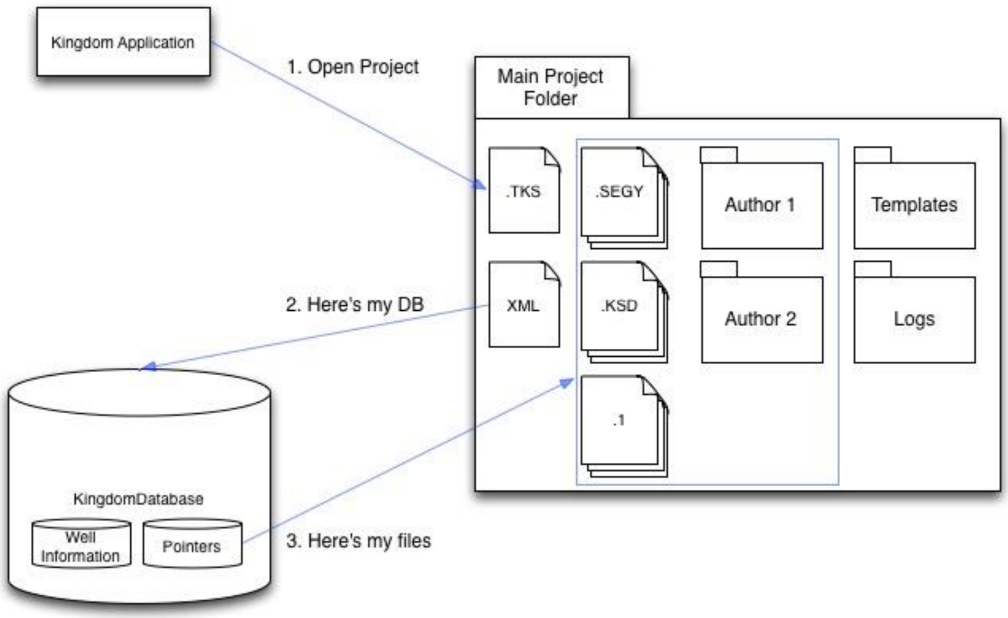
Each Kingdom 2015 project is stored in its own project folder or directory. All project files related to this particular project reside within this project folder by default and the files and subdirectories contained within this project folder.

These project folders can reside either on the users' workstation local drive where the Kingdom application runs or on network-attached storage seen as a Windows SMB share. Each project can be accessed by a single or multiple simultaneous users.

The files contained in the project folder consist mainly of seismic files and project files. There is a main project file that is unique to each project with the extension ".TKS." A project can be opened using the TKS file. There are several subdirectories, including ones for log files, templates, and authors. Each author (also known as an interpreter or user) has a private subdirectory created the first time the author opens a particular project. The author's interpretations, faults, and horizons are stored as files in this subdirectory within the project folder.

The seismic files can be fairly large files, whereas the other types of files are much smaller in size. The seismic files can be of different types: SEGY and KSD with their respective file extensions (.segy, .ksd, and .1). While SEGY is a standard input seismic file, other types are proprietary formats that are generated by the Kingdom application depending on the workflow an interpreter is performing. Upon generation, these files are automatically located in the main project folder along with the original seismic SEGY.

Figure 3) File and database layout.



2.2.2 Structured or Database Components

There is a component of Kingdom project data that resides in a database, as shown in Figure 1. The supported databases are SQL Express for small environments and SQL Server or Oracle for larger environments. Primarily the following types of information are stored in the database:

- Well information (majority of the DB)
- Location (pointers) of seismic files for each project and interpreter
- Location (pointers) of faults and horizons for each project and interpreter

By default, as mentioned previously, Kingdom places all newly generated files in the main project folder, and thus by default these database pointers just point to the same folder where the rest of the project data resides.

The well information, especially well logs, is by far the biggest data in the database in terms of size. Sometimes projects heaving in geology with a large DB take longer to open for a single user because all the relevant DB information is fetched to the user's workstation.

2.3 Traditional Backup Strategy

There are several backup applications that are used today for file-based backup of Kingdom main project data. This includes the seismic files that are typically stored within the main project folder. The Oracle database portion is typically backed up using database tools.

In case the project data is not all stored in one main folder, it is important to note that the project and all its components wherever they may reside have to come together after being restored for the project to open and function without errors.

The seismic files present a challenge for backups when they grow very large in size or quantity. Backups take too long, and data managers tend to relocate these files while updating the locations (pointers) in the database. In this scenario these relocated files need to be backed up or restored along with the project in a manual fashion. This makes the backup and restore more complex.

3 Understanding Kingdom Application Use Cases

In this section, we provide a high-level summary of some of the common use cases for Kingdom 2015. We have worked with IHS and some customers to select the following workflows/operations. The I/O characteristics of each of these workflows are as recorded on a NetApp Data ONTAP 8.3.1 platform. The test environment was set up similar to the one shown in Figure 1.

3.1 Convert Seismic File from SEGY Format to KSD Format

A seismic survey is like an ultrasound that is used to develop images of the rock layers below ground. This geophysical data is recorded in a special file format called the SEG-Y (SEGY). This format is the predominant industry standard format; however, it is not the most efficient and well-performing for use in the Kingdom application. Therefore, it is first converted to the KSD seismic file format, which is an IHS Kingdom proprietary format. This allows the seismic data to be reorganized in a way that is more space efficient and hence more well performing. Interpretation is then performed using the KSD seismic file. This workflow is also called “loading a SEGY.” It is a very common workflow, and therefore it is important to understand its I/O characteristics.

Note: The newly created KSD is automatically saved in the project folder.

Table 2) Converting seismic file from SEGY to KSD.

I/O	Metric
Read/write mix	87% read/11% write
Read pattern	100% sequential
Block sizes	35% 4k, 65% 64k for reads, 100% 64k for writes
Average IOPS load	1,400 SMB IOPS per user
File types	Seismic files
Cachability	Read from disk

Because the KSD format is a more compact format, we can see that the conversion workflow is a read-intensive workflow while it’s writing one-tenth of the time. The read operation is sequential as expected because the entire length of the SEGY file is read from beginning to end as it is converted. It is, however, a surprise to see a significant number of read requests at a 4k block size.

This is typically the first time in the project scope that the user reads the SEGY file, so it is read entirely from disk.

3.2 Impromptu Time Slice Generation

A time slice is a horizontal display of the 3D seismic data at a specific time from Kingdom’s VSD display. A time slice is a quick way to evaluate changes in the amplitude of seismic data. Kingdom can generate time slices immediately on the fly based on a particular crossline or inline display, and the information is held in memory only, not written to disk.

Table 3) Impromptu time slice generation.

I/O	Metric
Read/write mix	100% read
Read pattern	100% sequential
Block sizes	95% 64k and 5% 4k
Average IOPS load	1,500 (peaks at 3,000)
File types	Seismic files
Cachability	60% read from Flash Cache™

3.3 Generate Time Slice Volume

A time slice is a horizontal display of 3D seismic data at a specific time. A time slice is a quick way to evaluate changes in the amplitude of seismic data. Kingdom can generate time slices immediately based on a particular crossline or inline display, but to save time, a complete seismic 3D volume cube can be generated as well. In this workload, a complete time slice volume is generated and saved.

Table 4) Generating time slice volume.

I/O	Metric
Read/write mix	100% write
Block sizes	64k
Average IOPS load	1,500 (peaks at 3,000)
File types	Seismic files
Cachability	N/A

Large amounts of data are sent to the storage system and fill up 1Gbit link. However, we notice that very little is getting written to HDD. This is because a lot of the data being written to disk is actually zeroes, and with the zero block dedupe process in clustered Data ONTAP, it is able to eliminate these unnecessary writes to disk. This makes the workflow run more quickly and is a great advantage to the user.

3.4 Multiple Workflows

In this test environment, one workstation user is running crossline viewing (VSD vertical seismic display), while four workstation users are running TrackPAK generation of a seismic volume with attributes. This test incorporates multiple workflows.

Table 5) Multiple workflows.

I/O	Metric
Read/write mix	55% read/40% write
Read pattern	Sequential

I/O	Metric
Block sizes	15% 4k and 85% 64k for read, 64k for write
Average IOPS load	6,000
File types	Seismic files
Cachability	Read from disk

3.5 Crossline Viewing

Interpretation of the 3D seismic volume is performed by viewing the data in Kingdom's VSD of the inlines and crossline. As mentioned before, latency of this display getting from the disk to the screen is of big concern because it affects user experience. Viewing of crossline orientation tends to be slower than that of inline because data is being read from random locations of the seismic file.

Table 6) Crossline viewing.

I/O	Metric
Read/write mix	99% read
Read pattern	0% sequential
Block sizes	12k, 16k
Average IOPS load	500
File types	Seismic files
Cachability	Read from disk

3.6 Convert KSD Format to Brick Format for VuPAK

In bricked seismic data format, the Kingdom KSD file is grouped into three dimensional "bricks" of data, which allows users to optimize performance along crossline, inline, and time/depth scales. Kingdom's module viewer, VuPAK, allows the users to interpret the seismic volume in a 3D space. The bricked format increases the performance of viewing the data. This test characterizes the I/O for the process of converting a KSD file to brick format.

Table 7) Converting KSD to brick format for VuPAK.

I/O	Metric
Read/write mix	70–95% read
Read pattern	75–100% sequential
Block sizes	A mix of 4k and 64k for read, 64k for write
Average IOPS load	1,400
File types	Seismic files
Cachability	95% reads came from Flash Cache

Note: The KSD file was cached by Data ONTAP upon creation in section 3.3. Or perhaps "loaded brick volume" existed before this operation. Either way, the reads came from Flash Cache.

3.7 Illuminator Volume

Kingdom's illuminator volume is a process that creates an illuminator volume from a KSD seismic file. It is then used to increase the speed and accuracy of autopicking during horizon interpretation.

Table 8) Illuminator volume.

I/O	Metric
Read/write mix	75% read/22% write
Read pattern	80% sequential
Block sizes	16% 4k and 75% 64k for read 64k for write
Average IOPS load	600
File types	Seismic files
Cachability	Read from Flash Cache

3.8 Optimize Data Type Volume

This operation optimizes the 3D seismic data volume and creates a new data type that reorients the volume for increased loading and better performance.

Table 9) Optimize data type volume.

I/O	Metric
Read/write mix	21% read/75% write
Read pattern	75% sequential
Block sizes	13% 4k and 72% 64k for read 82% 4k and 17% 64k for write
Average IOPS load	1,200
File types	Seismic files
Cachability	Read from disk

3.9 TracePAK Volume

TracePAK is a process/module that is used to generate a new seismic volume based on various attributes of data needed for further interpretation.

Table 10) TracePAK volume/

I/O	Metric
Read/write mix	50% read/44% write
Read pattern	100% sequential
Block sizes	64k and smaller
Average IOPS load	1,200

I/O	Metric
File types	Seismic files
Cachability	Read from Flash Cache

4 NetApp Storage Recommendations

IHS Kingdom is a line of business application that facilitates core engineering related to exploration, discovery, and exploitation of trapped hydrocarbons. Needless to say, improvements in deployment, security and performance of this application and its related data are of utmost priority to the oil and gas company.

The primary users of this application are scientists, engineers, geophysicists, geologists, data quality engineers, field operators, and more. These users are collectively tasked to explore for hydrocarbons in the subsurface of the earth or deep in the ocean beds. For this purpose, the oil and gas companies go to great lengths and spend hundreds of millions of dollars to collect all sorts of data, including but not limited to seismic data, well log data, environmental data, and so on. This data is then put through various lengthy processes before it is consumed by these engineers. Therefore, this data is the most critical asset and the lifeblood of the exploration oil and gas company. It is of vital importance that this data be protected and backed up, the intellectual property created on the basis of this data be archived and protected, and in case of disasters this data be made available for business continuity through alternate means.

Moreover, the exploration oil and gas industry is highly competitive, much like any business, and to achieve that crucial competitive edge, the company needs to provide performance and accessibility of its most vital upstream data and applications to the end users.

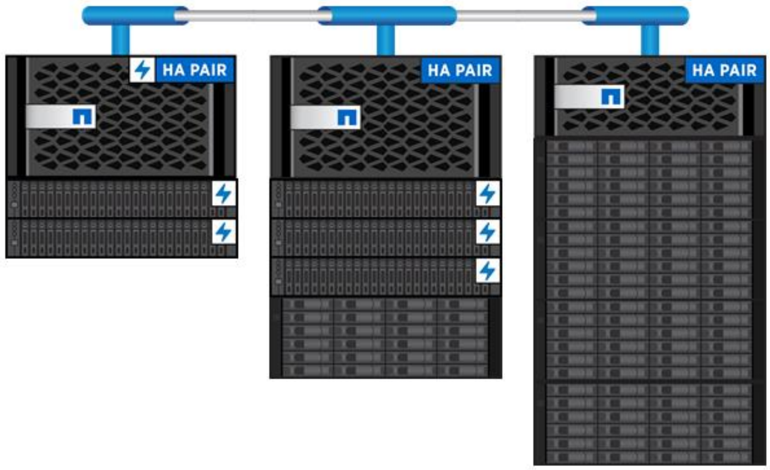
In summary, the project data associated with Kingdom needs to be highly available, with redundant copies allowing for disaster recovery and business continuity, and must be highly accessible to a diverse number, type, and geography of users. Moreover, the closed projects and historical seismic files need to be archived securely for retrieval years later when the respective fields are needed for further exploitation with newer technology.

In this section, our aim is to provide as much information as is necessary to design, size, and implement the best storage solution for the Kingdom application that meets the demanding needs of its users and helps them meet if not surpass their business objectives.

4.1 Clustered Data ONTAP

Clustered Data ONTAP is enterprise-capable, unified scale-out storage. It is the basis for virtualized, shared storage infrastructures. Clustered Data ONTAP is architected for the following features and functionality, all available in a single platform.

Figure 4) Data ONTAP cluster consisting of FAS controllers in a mix of all-flash, hybrid, and capacity configurations.



4.1.1 Nondisruptive Operations

Nondisruptive operations eliminate planned downtime. They enable you to continue serving data throughout your entire tech refresh lifecycle and to move workloads between all-flash, high-capacity, and hybrid tiers without reconfiguring your applications or affecting your users.

4.1.2 Storage and Operational Efficiency

Efficiency allows you to make the most of both data storage and your organization. Storage efficiency translates to a reduction in capacity requirements and more efficient use of flash. Cloning technologies also help you optimize your most important resource: the people who depend on infrastructure.

4.1.3 Scalability over the Lifetime of the System

Scalability allows organizations to reduce risk by knowing that the same tools and processes are available as needs grow and provides the flexibility to expand what is needed, when it is needed. Flash capacity can be increased for performance, and high-density drive capacity can be increased for archive data. Clustered Data ONTAP allows you to scale up when you need a higher-end storage array or scale out horizontally when you need to distribute a workload. You can scale compute in the cloud while maintaining control of your data.

4.2 Recommendations for Performance

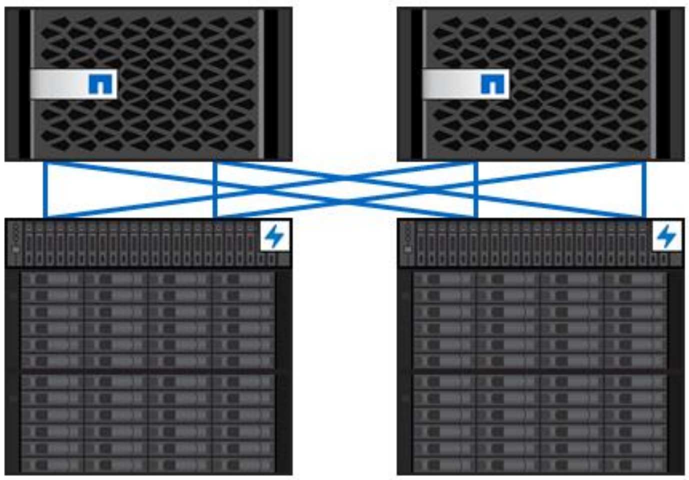
Performance of the data is very important for oil and gas exploration and production applications. There are mainly two types of applications that need to be considered for performance: compute based and user driven. Compute-based applications such as seismic processing are simpler, where the data is typically accessed in a uniform fashion, albeit at higher throughput. In contrast, user-driven applications such as interpretation can be very complex. These applications access nonuniform data types from different sources. They simultaneously rely on structured and unstructured data. They also have complex security, backup, archiving, and DR strategies. Therefore, they have high performance requirements as well as complex data management needs. They simultaneously have continuous, reliable, and high-performance access to data.

Designing a comprehensive performance storage strategy for upstream oil and gas data requires fully leveraging the advanced features of clustered Data ONTAP. We discuss some of the main characteristics of the design in the following sections.

4.2.1 Disks, Aggregates, Volumes, and Flash Cache

Clustered Data ONTAP supports various types of disk medium. It does this while providing the same advanced features to the users and applications and without changing the way they access or store data to the different disk mediums.

Figure 5) FAS8000 controller showing one shelf each of solid state disks and a few shelves of capacity drives.



In a clustered Data ONTAP FAS system, the physical drives managed by nodes are grouped together in RAID groups. We recommend RAID DP® for configurations supporting the Kingdom application data. RAID groups are grouped together to form an aggregate. An aggregate can consist of solid-state drives (SSDs), spinning media, or both. An aggregate is the large pool of storage that we can use to carve out usable chunks of storage called volumes. The number of drives in all the RAID groups within the aggregate should be the same.

An aggregate that consists of both SSDs and capacity drives is called a Flash Pool™ aggregate. The solid-state drives allow read and write caching of data just for that aggregate. It is worth mentioning now that in contrast, Flash Cache allows read caching of data from all the aggregates in the system except Flash Pool aggregates and SSD aggregates.

Clustered Data ONTAP 8.3 provides advanced drive partitioning (ADP), which allows single physical disks to be shared by multiple aggregates. NetApp supports ADP for All Flash FAS, entry platforms, and NetApp Flash Pool configurations.

4.2.2 Optimal File System Layout

The Kingdom application data is composed of various data types, as shown in Figure 2. The important thing to note is that each data type has a different I/O pattern. We conducted several tests to figure out the I/O pattern of each data type or dataset by examining the associated workloads from a user perspective.

Our results show that a storage system that is a hybrid of disk technologies provides the perfect balance between price and performance for IHS Kingdom. Meanwhile, All Flash FAS can offer consistent best performance for all types of workflows, keeping the data management and storage efficiency features in addition to the availability and scalability intact.

We recommend storing the seismic SEGY data in a separate volume (and node) on a SATA aggregate. The SEGY is read once and never updated. It is also not cached and very large. Only read performance is needed, and because it's always read sequentially, it makes the most sense to store it on SATA.

The KSD and brick seismic data should reside on a Flash Pool aggregate with SSD and SAS (or SATA when density is of higher concern than performance). This is because KSD is in use all the time for interpretation, and the brick format is used for visualization in geology. These files are smaller than SEGY, but one key difference other than size is that these can be cached because they are not always being read sequentially. Therefore, Flash Pool aggregates provide the best value for storing KSD and brick seismic files while providing the required performance through solid-state caching. The cache policy on the Flash Pool aggregate should be set to read because these files don't change after being created.

The rest of the project data should reside in a volume on a SAS aggregate with Flash Cache or Flash Pool. The choice between Flash Cache and Flash Pool depends on the size of the active working dataset and whether the aggregate is shared for KSD as well as project data. The benefit of Flash Pool over Flash Cache is not just the larger read cache but also the ability of cache to persist across reboots.

The recommendation for the Oracle database is dependent on the size of the database and the user community. For a large environment, it is recommended to store the database in a volume on a Flash Pool aggregate with SAS capacity drives. For a smaller environment, SAS aggregates complemented with Flash Cache would be sufficient.

Last, it is important to point out that as projects grow in size and several to many simultaneous users access the same very large projects, keeping the latency down becomes a big challenge, especially for the file access portion of the dataset. In these extreme cases, it is recommended to store the project data on All Flash FAS. An All Flash FAS is a FAS with all-SSD shelves and special flash optimizations to deliver consistent low-latency I/O to the Kingdom application.

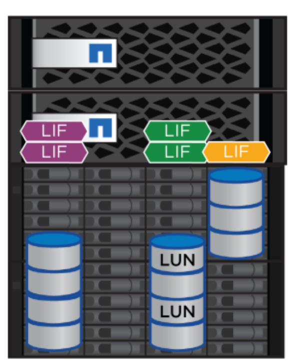
Clustered Data ONTAP clusters allow all the different types of nodes and configurations described earlier to be included in the same cluster. We discuss some of the other features and strategies in the following sections.

4.2.3 Storage Virtual Machine, Unified Namespace, and Networking

A cluster provides hardware resources, but clients and hosts access storage in clustered Data ONTAP through storage virtual machines (SVMs). SVMs exist natively inside of clustered Data ONTAP. SVMs define authentication, network access to the storage in the form of logical interfaces (LIFs), and the storage itself in the form of SAN LUNs or NAS volumes. The clients and hosts connect to an SVM, rather than to a physical storage array.

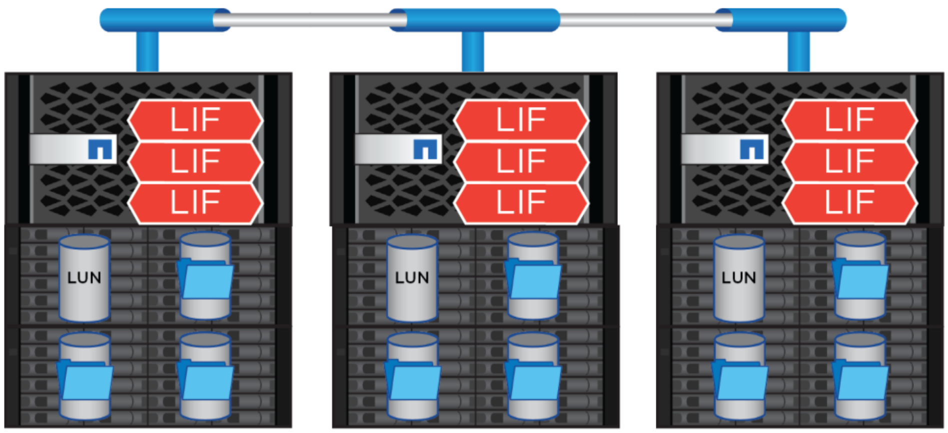
Like compute virtual machines, SVMs decouple services from hardware. Unlike compute virtual machines, a single SVM may use the network ports and storage of many controllers, enabling scale-out. One controller's physical network ports and physical storage may be also shared by many SVMs, enabling multitenancy. Figure 6 shows a small cluster with three SVMs. Each SVM has its own volumes, LUNs, network connectivity (LIFs), and authentication. Figure 7 shows a six-node cluster with one SVM. Clients can access data from any node and any network LIF.

Figure 6) Small cluster with three SVMs.



The components of an SVM are not permanently tied to any specific piece of hardware in the cluster. An SVM's volumes, LUNs, and logical interfaces can move to different physical locations inside the cluster, while maintaining the same logical location to clients and hosts. While physical storage and network access move to a new location inside the cluster, clients can continue accessing data in those volumes or LUNs, using those logical interfaces.

Figure 7) Six-node cluster with one SVM.



The Kingdom application clients are typically connected to the data center over 1Gbit Ethernet, as shown in Figure 1. This limits the ingest of large seismic files to a maximum of 120MBps on SMB. The SEGY is typically ingested when converting to KDS format. During this process, we observe that a 1Gbit link can be over 80% utilized. This indicates that an upgraded link to 10Gbit may help with this workflow. However, the question to ask is whether this justifies the cost of upgrading every Kingdom user to 10Gbit Ethernet. Each customer really has to weigh the benefit to this single use case compared to the costs themselves. In our testing, we concluded that because most of the daily workflows discussed in the use cases section earlier are more compute intensive and can be accommodated in a 1Gbit link per workstation, the focus should be on lowering latencies instead for the majority cases.

For Kingdom users, as they access a single SMB share for the entire project on one node, while the seismic data may actually be stored on a different node, it is ideal that they access the share on a LIF that resides on the node with the KSD and brick seismic data on it. This is because the seismic SEGY is only accessed infrequently, and the other types of seismic data (that is, KSD and brick) are read quite frequently and in a random fashion, making it much more important to reduce the latency. The project data should be on the same node as the other seismic data in most cases.

4.3 Efficiency of the Data

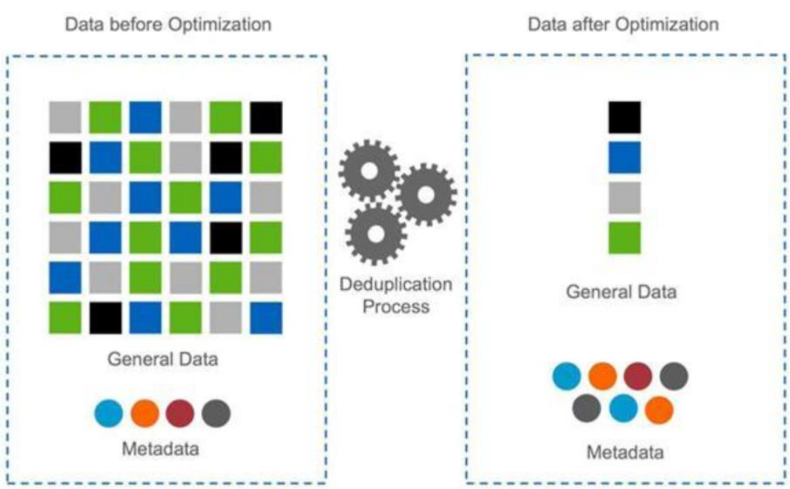
Storage efficiency built into clustered Data ONTAP offers substantial space savings, allowing more data to be stored at lower cost. For the Kingdom application, we are going to be considering volume data deduplication at a 4k block level and volume-level compression because these are of great interest to O&G customers and Kingdom users. NetApp deduplication and data compression are two key components of NetApp storage efficiency technologies that enable users to store the maximum amount of data for the lowest possible cost. Other efficiency features include but are not limited to thin provisioning, FlexClone® technology, and virtual storage tiering.

4.3.1 Deduplication

Part of NetApp's storage efficiency offerings, NetApp deduplication provides block-level deduplication within a flexible volume. Essentially, this process removes duplicate blocks, storing only unique blocks in the volume, and it creates a small amount of additional metadata in the process. Notable features of this process include:

- It works with a high degree of granularity: at the 4KB block level.
- It operates on the active file system of the volume. Any blocks referenced by Snapshot® copies are not made available to deduplication.
- It is a background process that can be configured to run automatically on a schedule or run manually.
- It is application transparent.

Figure 8) How NetApp deduplication works at a high level.



As we learned in the previous sections, a Kingdom project consists of different types of files. Typically, they all reside in the project folder. In larger environments, it's recommended to break out the seismic files from the project data for optimized performance. Our tests indicate that the seismic files are typically not a good candidate for deduplication except in a special case in which there are several copies of the same seismic file in the same volume. This scenario is highly unlikely. Therefore, we conclude that deduplication does not yield any significant results when run on a volume with only SEGY seismic files. However, our tests showed interesting results for KSD and brick formats that we share in the following sections.

The project data has different characteristics than the large seismic files. Deduplication still helps save storage capacity for the project data (nonseismic) files portion of the project. Our tests on a small volume with project data have shown about 30% space savings from deduplication alone on entire project folders.

4.3.2 Data Compression

NetApp data compression is a software-based solution that provides transparent data compression. It can be run inline or postprocess and also includes the ability to perform compression of existing data. No application changes are required to use NetApp data compression.

Postprocess compression uses the same schedule as deduplication. If compression is enabled when the volume efficiency schedule initiates a postprocess operation, it runs compression first, followed by deduplication. It includes the ability to compress data that existed on disk prior to enabling compression.

NetApp data compression can be configured as an inline operation. In this way, as data is sent to the storage system, it is compressed in memory before being written to the disk. The advantage of this implementation is that it can reduce the amount of write I/O. This implementation option can affect your write performance and thus should not be used for performance-sensitive environments without first doing proper testing to understand the impact.

In order to provide the fastest throughput, inline compression compresses most new writes but defers some more performance-intensive compression operations to compress when the next postprocess compression process is run. An example of a performance-intensive compression operation is a small (<32k) partial file overwrite.

If both inline and postprocess compression are enabled, then postprocess compression tries to compress only blocks that are not already compressed. This includes blocks that were bypassed by inline compression such as small partial file overwrites.

Kingdom seismic files are compressible. For our compression tests, we collected some real seismic data from various Kingdom projects at IHS. The tests were conducted at an IHS facility. The files were stored in individual volumes and populated with lots of SEGY, KSD, and brick format files. The volume containing SEGY seismic files in our tests compressed at an impressive 50%. The volume containing KSD seismic files, which is a Kingdom proprietary format used for interpretation, compressed at 39% but with total efficiency savings of 46% (including deduplication). Last, the third volume containing the type of seismic file, which is the brick seismic file format used by VuPAK for displaying 3D seismic cubes on user monitors, was compressed down to 29%. These results are not surprising because the brick format is smaller and somewhat already optimized on a file system. What is interesting is that the brick format in addition to compression also yielded deduplication savings of about 33%. The other two seismic file formats yielded less dedupe savings, as expected.

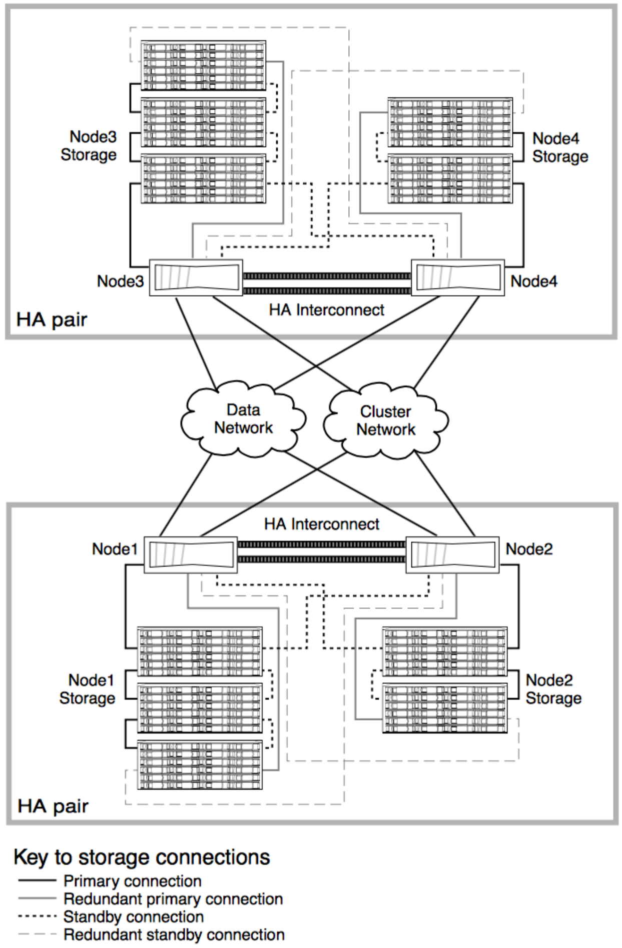
The bottom line is that the NetApp storage efficiency features, especially deduplication and data compression, are effective in providing space efficiency for Kingdom project data as well as seismic volumes.

4.4 Availability of the Data

4.4.1 High Availability and Scale-Out

Clustered Data ONTAP FAS systems are designed for very high availability. NetApp FAS technology has an uptime of six 9s or higher across all monitored systems. This means extreme scalability and a rich feature set to improve performance and manageability while providing extreme availability. This is why clustered Data ONTAP is one of the leading storage operating systems in the upstream oil and gas industry for applications such as Kingdom.

Figure 9) Two HA pairs consisting of 4 nodes in a cluster.



A high-availability Data ONTAP cluster consists of two nodes as a unit. These nodes in the HA pair can take over each other's storage in the event that one of the nodes goes down due to some failure.

The controllers in the cluster are also connected to each other using a network called the cluster network. This network facilitates intercluster communication, allowing for applications to access data on any disk from any node, regardless of to which node the disks are physically connected. This architecture allows for not only extreme high availability but also tremendous scale-out.

Scale-out means adding capacity or compute power to an existing cluster without any disruption to the ongoing service. New nodes are added, and the additional resources are made available immediately. Similarly, leveraging the clustering technology, old controllers can be removed and replaced by new ones with a bit of data management (that is, relocating volumes off the old disks and onto new disks). This process of recycling old gear is also done nondisruptively.

NetApp recommends distributing Kingdom project and seismic data across a cluster in more than one node, primarily the project data volumes, KDS, and brick volumes on one node and the SEGY seismic files on another node.

4.4.2 Backup, Disaster Recovery, and Business Continuity

Backing up Kingdom projects can be very simple to extremely complicated, based on the size of the project. As discussed in a previous section, a Kingdom project is composed of many different types and sizes of files. These files are linked to the project through a database (Oracle or SQL). This allows the data managers to relocate the seismic files to other locations on different network shares outside the main project folder.

This strategy solves one data management problem but introduces another. Backups and restores are now smaller and complete in less time because we are not backing up the seismic files (that never change) in every backup job. However, the projects now don't contain all the files that are part of the project in the main folder, and backups are missing critical parts of the project unless those files are located and backed up along with the project. There can be a few different ways to get this done right, but it's a manual process today. Hence, a little more data management is involved to achieve better performance for larger environments. After being restored, the missing seismic data needs to be relinked with the project.

We know that the seismic files, once created, don't change during their lifetime. They are a read-only workload after creation. Therefore, doing a full backup of these files each time the project backup is done isn't only redundant, it's also taking up a lot of space on the backup platform.

5 Sizing Storage for Kingdom Users Using SPM

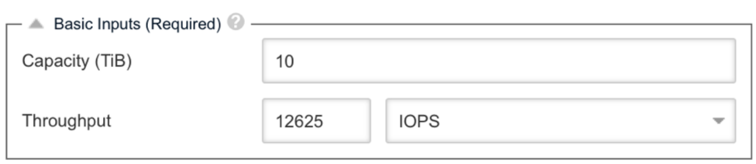
5.1 Capacity

Kingdom environments can be rather large in terms of capacity. However, it's relatively simple to size for capacity. Customers in this industry have a good idea of how much seismic data they have in existing storage and how much they expect to receive in the near future. They have specific requirements in terms of terabytes of how much they would like to expand or purchase. Their requirements are almost always in usable capacity. The exceptions are in some large organizations where the data storage group has already translated the user requirements to raw capacity. In either case, it's important to clarify.

The project data tends to grow as well, but not in big chunks like seismic data. Projects grow organically through user interactions with their Kingdom projects. The growth is mostly attributed to the new proprietary seismic files being created (for example, KSD and brick from imported SEGY). Just like before, this project data growth estimate should be calculated by the customer/user.

After the total capacity is determined, the data is laid out in the file system according to the file structure explained in section 2.2 of this document.

Figure 10) Input capacity value and select IOPS in NetApp System Performance Modeler.



5.2 Performance

Data managers and application end users are always looking for better performance of their workloads. There is an especially strong relationship between application performance, data access performance, and user productivity in the upstream oil and gas workflows. In order to size the best performing clustered Data ONTAP system for a given Kingdom environment, we first try to determine typical, average user workload I/O characteristics. We incorporate all the use cases we have selected for a typical user in section 3. Sizing for the database portion of Kingdom is outside the scope of this document, and some guidance can be obtained from the Oracle Quick Reference Guide. A link to this QRG can be obtained from the references section. Furthermore, we use weighted averages to determine the level of concurrency to make sure we don't end up oversizing.

Note: The information provided here for the purpose of sizing is meant to be used in the NetApp System Performance Modeler (SPM) tool.

The numbers in the following series of SPM diagrams reflect the outcome of assigning the appropriate weighted averages to all the use cases in section 3 and averaging them out.

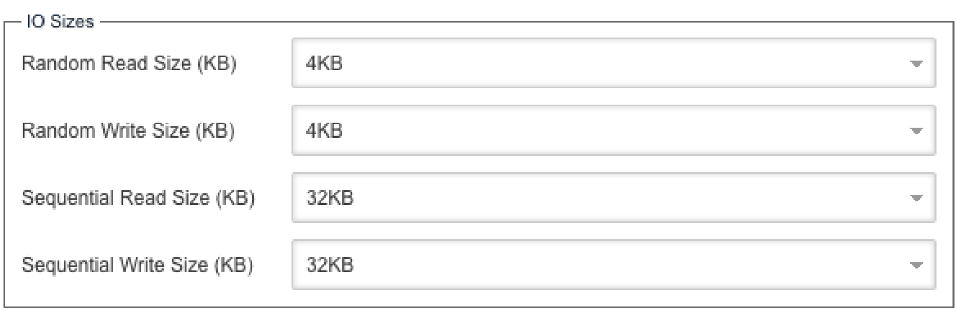
First, we need to select the appropriate number of IOPS for the desired system. Our weighted average calculations have revealed a per user IOPS load of 1,265. Using this, we simply multiply by the number of total users the customer wants to support with good performance on this system. In Figure 9 we can see an estimate for 10 users.

Figure 11 shows the read and write as well as the sequential and random I/O breakdown.

Figure 11) Read and write breakdown for an average weighted Kingdom user.

SPM does not accept a block size higher than 32KB for sequential I/O, so we go with that instead of 64KB. For our sizing purposes, this does the job.

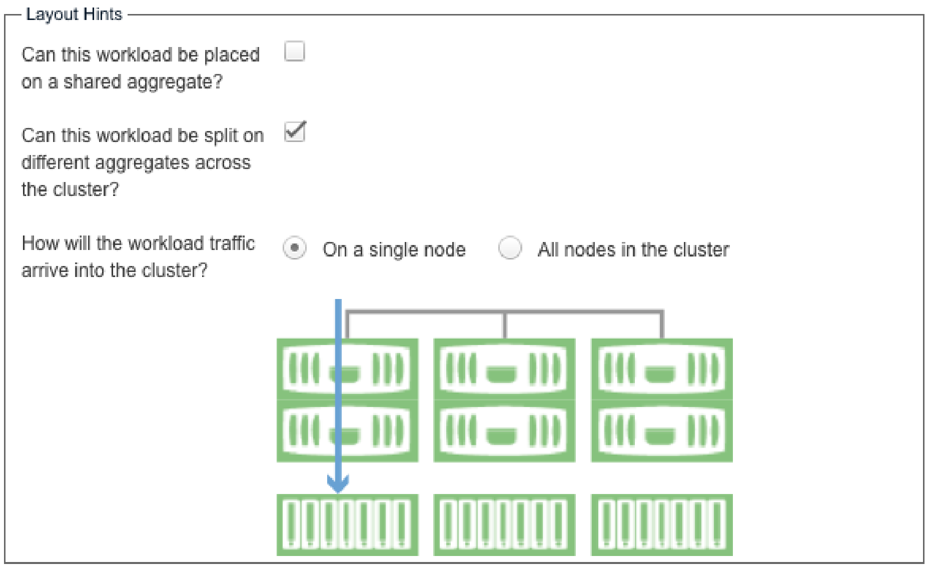
Figure 12) Block sizes for Kingdom.



The CIFS option should be selected because there is no NFS workload for Kingdom. However, if the system is being designed for multiple workloads in addition to Kingdom, then NFS may be selected for other workloads. In the majority of customer environments, all the Kingdom users access the same network share for their projects. This means a single host name or IP is provided as a mapped drive on their workstations. For this simple approach, it is recommended to select "single node" for traffic access (in Figure 13) to the cluster. However, for advanced user environments where they have a data or storage management team and who are familiar with the application/user environment, a better approach may be to select "All nodes in the cluster." In this case several IPs (LIFs) are associated with the share, and different groups of users are mapped through those separate IPs.

In larger environments, it is recommended to use different aggregates for seismic and project data, as outlined in the file layout section of this document.

Figure 13) Access to the cluster from the workstation.



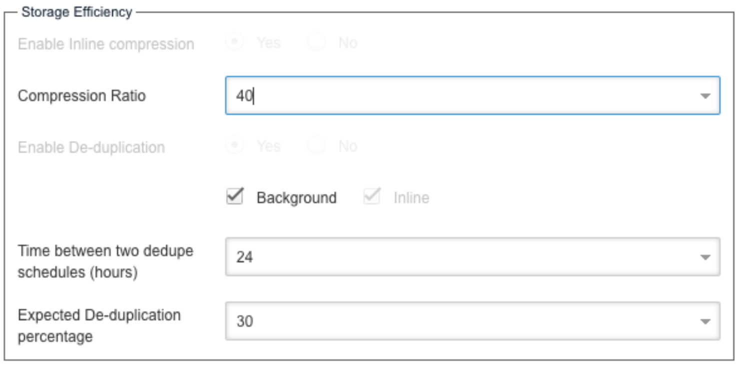
The active working set size for these users is going to be on the larger side. The main reason for this is the cacheability of the brick format seismic file. These files are smaller than the huge SEGY files, but they can still be in the 100s of gigabytes range. Depending on customer requirements, this value should be set anywhere from 10% to 20%.

Figure 14) Active working set.

The random read latency, which affects mostly the crossline viewing use case, should be set to 10ms or lower if this use case is a big concern to the customer. Normally, this is the slowest operation and requires the users to wait. It's really a customer preference if their users want to view crossline on the fly or convert the seismic file to be optimized for crossline. Both involve waiting; however, in the latter procedure, the viewing is no longer random.

Last, the storage efficiency estimates should be discussed with the customer first. They can either size the system with incorporating the estimates or size without them and just use the extra storage as additional capacity. Note that in the cases where customers are migrating from an older storage system, they may need the full capacity to begin with before the efficiency savings kick in.

Figure 15) Deduplication and compression estimates.



After these numbers are entered into the SPM, several possible configurations are provided. At this point the configurations need to be checked for other criteria outside of IHS Kingdom (for example, network connectivity requirements, additional workloads, multisite, backup, and DR).

6 Conclusion

IHS Kingdom is a powerful seismic interpretation platform and is widely deployed throughout the world in the oil and gas exploration and extraction industry. The application covers a diverse set of workflows that apply to various substages of exploration and involve different datasets.

Clustered Data ONTAP not only meets and exceed the tremendous requirements of growth, scale, and uptime for the IHS Kingdom application and the oil and gas industry, as we have seen from our testing. It also provides excellent efficiency and performance. As macroeconomic factors have a huge impact on this industry relative to supply and demand and the price of oil, the large and small players really see the efficiency and performance as critical requirements.

References

This report references the following documents and resources:

- Kingdom 2015 Release Notes, July 2014: http://www.seismicmicro.com/ProductReleaseDocumentation/KINGDOM/ReadMe.pdf
- Two Alternative Seismic Fault Interpretation Techniques by Mike Cline: http://www.hgs.org/node/4306
- NetApp TR-3982: NetApp Clustered Data ONTAP 8.3.x and 8.2.x: An Introduction: http://www.netapp.com/us/media/tr-3982.pdf
- NetApp TR-3966: NetApp Data Compression and Deduplication Deployment and Implementation Guide: http://www.netapp.com/us/media/tr-3966.pdf

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

NetApp
www.netapp.com