IBM Spectrum Scale and ECM FileNet Content Manager Are a Winning Combination
Deployment Variations and Value-added Features

Sandeep R. Patil
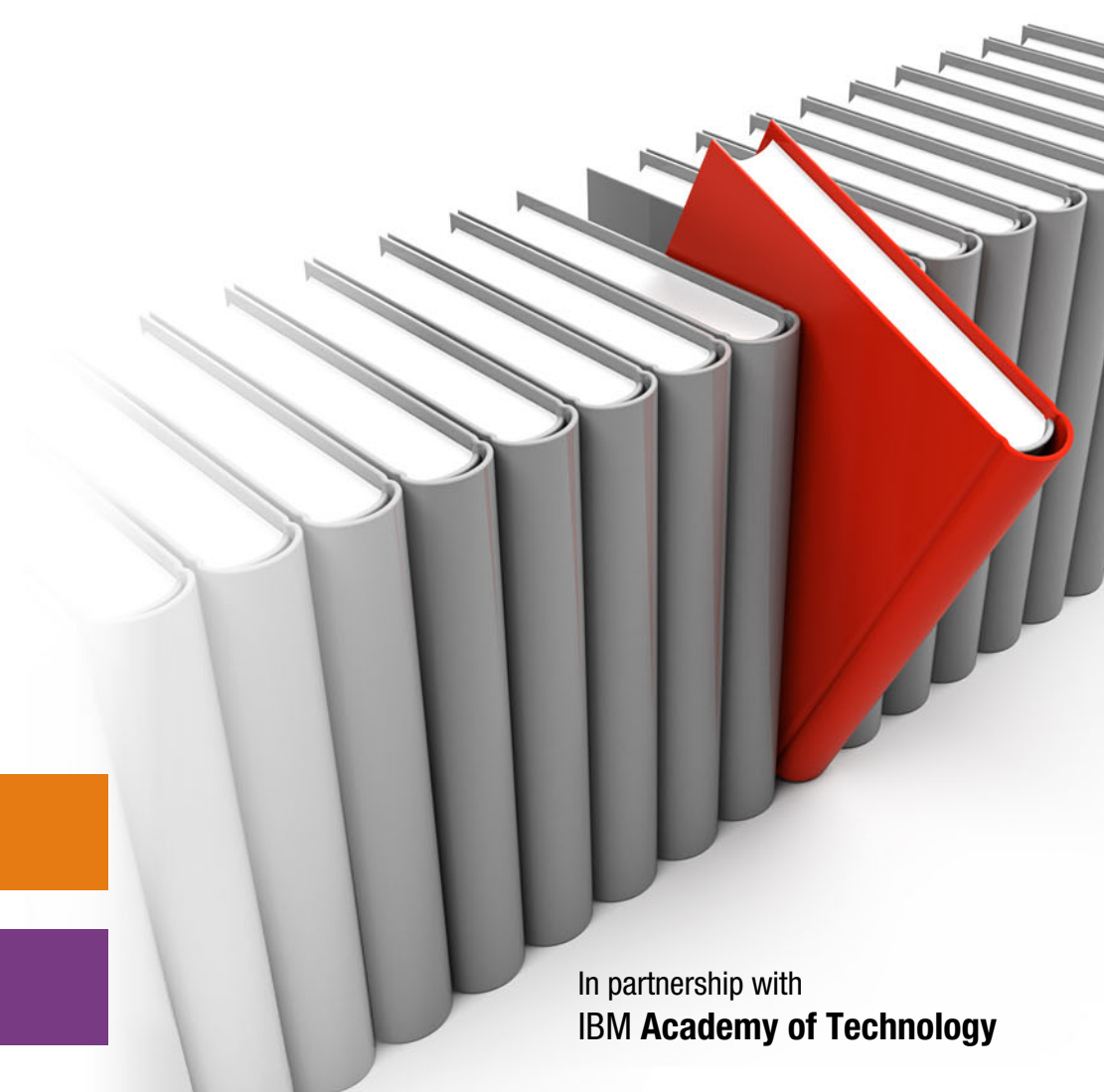
Atul Gore

Sasikanth Eda

Michael Bordash

Sanjay K. Sudam

Sathish Subramanyam

Analytics

Storage

In partnership with
IBM **Academy of Technology**

Redpaper

Find and read thousands of IBM Redbooks publications

- ▶ Search, bookmark, save and organize favorites
- ▶ Get personalized notifications of new content
- ▶ Link to the latest Redbooks blogs and videos

Get the latest version of the Redbooks Mobile App

iOS

Download Now

Android

Promote your business in an IBM Redbooks publication

Place a Sponsorship Promotion in an IBM® Redbooks® publication, featuring your business or solution with a link to your web site.

Qualified IBM Business Partners may place a full page promotion in the most popular Redbooks publications. Imagine the power of being seen by users who download millions of Redbooks publications each year!

It's good to be noticed.

ibm.com/Redbooks
About Redbooks → Business Partner Programs

THIS PAGE INTENTIONALLY LEFT BLANK

Abstract

This IBM® Redpaper™ publication describes various deployment options to configure IBM enterprise content management (ECM) FileNet® Content Manager components to use IBM Spectrum Scale™ (formerly known as IBM GPFS™) as back-end storage. The paper also describes various IBM Spectrum Scale value-added features with FileNet Content Manager to facilitate an efficient and effective data-management solution.

Prerequisites

This paper assumes that you are familiar with the following requisites:

► Basic IBM Spectrum Scale implementation and administration.

► IBM Enterprise Content Management concepts and administration.

Introduction to IBM Spectrum Scale

IBM Spectrum Scale is a proven, scalable, high-performance data and file management solution. It provides world-class storage management with extreme scalability, flash accelerated performance, and automatic policy-based storage that has tiers of flash through disk to tape. IBM Spectrum Scale reduces storage costs up to 90% while improving security and management efficiency in cloud, big data, and analytics environments.

IBM Spectrum Scale Version 4.2 provides highly differentiated value:

► Virtually limitless scaling to nine quintillion files and yottabytes of data.

► High performance, over 400 GBps, and simultaneous access to a common set of shared data.

► Integrated information lifecycle management (ILM) tools to automatically move data based on policies. This can dramatically reduce operational costs as fewer administrators can manage larger storage infrastructures.

► Software-defined storage so you can build your infrastructure your way:

   – Easy to scale with relatively inexpensive commodity hardware while maintaining world-class storage management capabilities.

   – Supports any combination of flash, spinning disk, and tape.

   – Supports various cluster models that include storage area networks (SANs), Network Shared Disk, and shared nothing clusters.

   – Supports addition of more storage capacity, without affecting the application, to greatly simplify administration.

   – Cross-platform solution so you can work with different platforms and operating systems. You can create an IBM Spectrum Scale cluster using IBM AIX®, Linux, and Windows server nodes, or a mix of all three. IBM Spectrum Scale is also available for IBM z Systems™.

► Global data access across geographic distances and unreliable WAN connections.

► Proven reliability with use in the most demanding commercial applications.

► Protects data from most security breaches, unauthorized access, or being lost, stolen, or improperly discarded with native file encryption for data at rest and secure erase.

► Multi-site support, connecting local IBM Spectrum Scale cluster to remote clusters to provide disaster recovery configurations.

► Offers additional protocol access methods in the Standard and Advanced editions of the product. Providing these additional file and object access methods and integrating them with IBM Spectrum Scale offers several benefits:

   – Users can consolidate various sources of data efficiently in one global namespace.

   – Offers a unified data management solution and enables efficient space utilization but avoids having to make unnecessary data moves just because access methods might differ.

The additional protocol access methods integrated with IBM Spectrum Scale are file access using Network File System (NFS) and Server Message Block (SMB), and object access using OpenStack Swift. Although each of these server functions (NFS, SMB, and object) uses open source technologies, this integration adds value by providing the ability to scale and by providing high availability using the clustering technology in IBM Spectrum Scale

- Provides policy-driven compression to reduce the size of data at rest.
- The GUI simplifies many storage administrator tasks and delivers monitoring capabilities.
- Has block storage for OpenStack deployments through its cinder driver support.
- Is a POSIX-compliant file system that offers an enterprise-class alternative to Hadoop Distributed File System (HDFS) and is a preferred platform for data analytics where its Hadoop compatibility extensions help replace HDFS in a Hadoop ecosystem, with no changes required to Hadoop applications
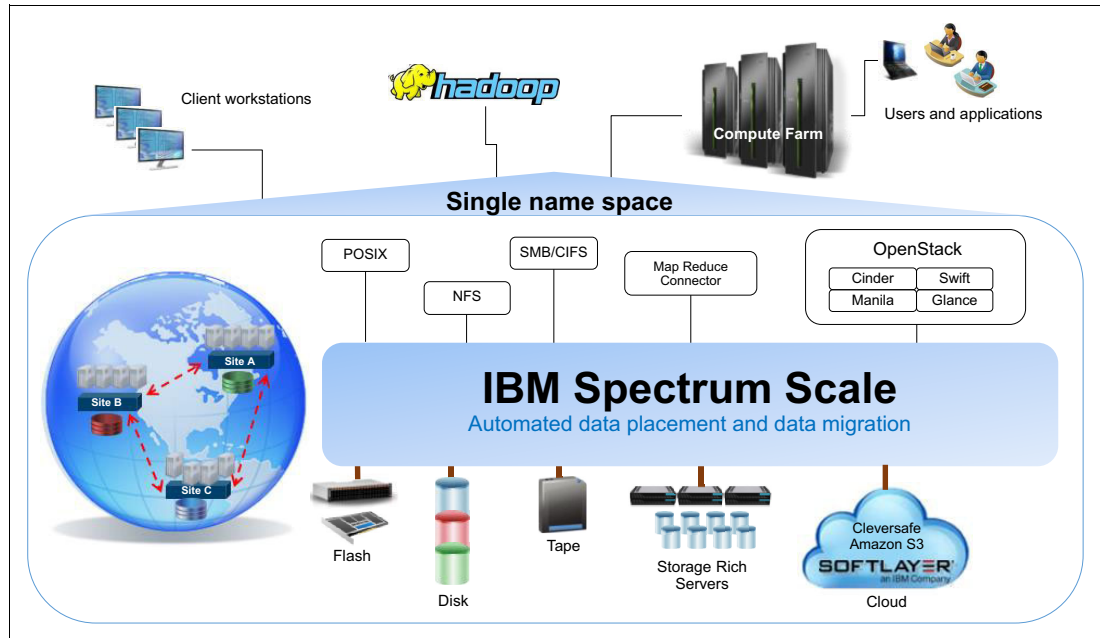
Figure 1shows an overview of IBM Spectrum Scale.

Figure 1   IBM Spectrum Scale overview

IBM Spectrum Scale powers many of the world's largest scientific supercomputers and commercial applications that require high-speed access to large volumes of data, such as digital media, engineering design, business intelligence, financial analytics, seismic data processing, geographic information systems, scalable file serving, and others.

Note: IBM Spectrum Scale offers a no-cost try and buy IBM Spectrum Scale Trial VM. The Trial VM offers a fully preconfigured IBM Spectrum Scale instance in a virtual machine, based on IBM Spectrum Scale 4.2 GA version. You can download it from this web page:

https://www.ibm.com/developerworks/servicemanagement/tc/gpfs/evaluate.html

Introduction to IBM enterprise content management

IBM enterprise content management (ECM) high-value solutions can help companies transform the way they do business by enabling companies to put content in motion by capturing, activating, socializing, analyzing, and governing it throughout the entire lifecycle.

ECM solutions help companies and government agencies reduce costs and make quick, smart, and cost-effective decisions. ECM solutions can be organized into these areas:

Essential ECM

Manages essential content anywhere. Production imaging and capture, enterprise report management, office document management (ODM), and standardization and consolidation are content management foundations for reducing costs and improving efficiencies.

Advanced case management

Takes workflow and business process management to the next level. It brings people, processes, and information together to drive better case outcomes.

Information lifecycle governance

Expands the value of repositories. It can reduce costs and risks while improving compliance posture.

Trusted content analytics

Derives unexpected content insights.

ECM solutions take care of content, security, storage, workflows, decision-making, and productivity. They also support rapid application development, using Web 2.0 technology in user-oriented application environments. That accelerates the creation of solutions and reduces dependence on IT resources.

FileNet Content Manager Platform ECM solutions

The IBM FileNet Content Manager Platform provides a breadth and depth of core functionality, enabling enterprise solutions. FileNet Content Manager provides content, security, and storage. FileNet Business Process Manager supplies workflows, decision-making, and productivity. These two core solutions facilitate resolving critical organization needs and optimizing business outcomes.

FileNet Content Manager helps organizations optimize processes, shorten production times, and improve productivity and accuracy. It includes process design and simulation tools, electronic forms, application development frameworks, and monitoring and reporting tools. It also includes a user interface designed for the business user. FileNet Content Manager also supports collaboration, widgets, integration with Microsoft Office 2016, and more.

The platform also supports rapid application development, leveraging Web 2.0 technology in user-oriented application environments. This accelerates creation and deployment of case-based applications across the enterprise. FileNet includes pre-wired ECM Widgets in a Web 2.0 Mashup-based framework that accelerates creating solutions and reduces dependence on IT resources. By using open standards, corporations harness maximum flexibility and interoperability with popular third-party applications and frameworks.

For more information, see the following web page:

http://www.ibm.com/software/products/en/fncm

Deployment topologies of ECM FileNet with IBM Spectrum Scale

IBM Spectrum Scale is based on software-defined storage principals and provides various cluster topologies that are described in the IBM Redbooks® publication, *IBM Spectrum Scale (formerly GPFS)*, SG24-8254. This paper describes ECM - IBM Spectrum Scale deployment topologies based on IBM Spectrum Scale cluster offering through IBM Elastic Storage™ Server. However, the content is applicable to other IBM Spectrum Scale cluster topologies.

This section illustrates the various ECM FileNet Content Manager deployment topologies that an administrator can use to leverage the access protocols offered by IBM Spectrum Scale. Of the deployment topologies that are described, which one to select depends upon the overall use case available.

> **Notes:**
>
> ► For the IBM Spectrum Scale supported clustering configuration, see Chapter 2 of *IBM Spectrum Scale (formerly GPFS)*, SG24-8254:
>
>   http://www.redbooks.ibm.com/abstracts/sg248254.html
>
> ► For details about IBM Elastic Storage Server, see *Introduction Guide to the IBM Elastic Storage Server*, REDP-5253:
>
>   http://www.redbooks.ibm.com/abstracts/redp5253.html

IBM Spectrum Scale POSIX interface: ECM configuration

IBM Spectrum Scale offers a rich POSIX interface to consumers and to applications. Figure 2 shows a high-level deployment diagram representing a basic FileNet Content Manager platform (including Content Platform Engine, Application Engine, Database: IBM DB2®) configured to use IBM Spectrum Scale POSIX interface.
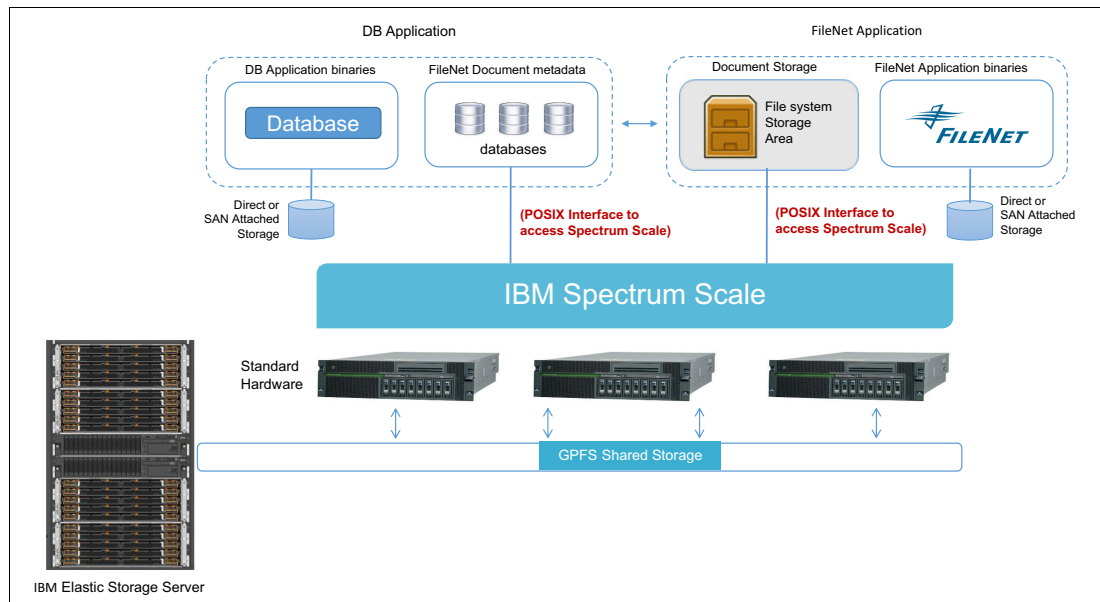

Figure 2   FileNet Content Manager components using IBM Spectrum Scale POSIX interface

In Figure 2, note that IBM Spectrum Scale is installed on standard hardware (Intel or AMD Opteron based servers, P-series, and IBM z Systems) and is supported on all major operating systems (AIX, Windows, RHEL, Ubuntu, and SUSE). Large IBM Spectrum Scale clusters are typically built by using two kinds of licenses: server and client. For more details related to license differentiation, see the following web page in the IBM Knowledge Center:

https://ibm.biz/Bd4Bgx

However, this topology supports use of both licenses and combination.

This deployment is generally the most preferred because the POSIX interface gives direct connectivity to the file system and proves to be performance-savvy.

IBM Spectrum Scale SMB, NFS interface: ECM configuration

IBM Spectrum Scale offers a full-fledged SMB, NFS interface to consumers and to file-serving applications. Figure 3 is a high-level deployment diagram representing a basic FileNet Content Manager platform (including Content Platform Engine, Application Engine, Database (DB2)) using IBM Spectrum Scale SMB, NFS interface.
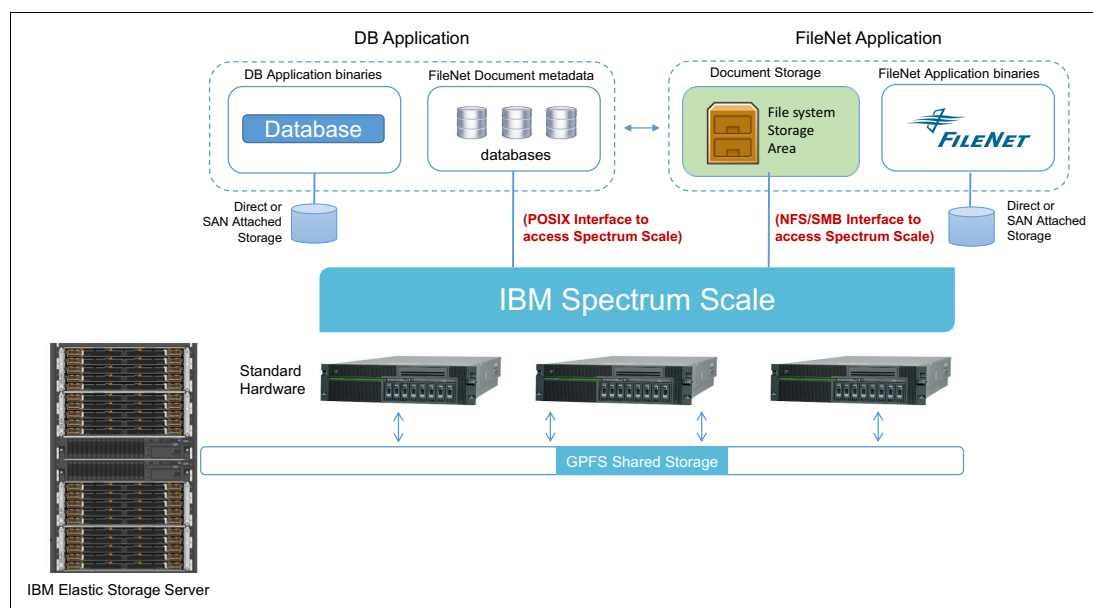

Figure 3   FileNet Content Manager components using IBM Spectrum Scale SMB/NFS interface

In Figure 3, note that the FileNet Content Manager database component is installed, configured using POSIX interface, and that the FileNet storage area (advanced file system storage device) is configured to use a SMB/NFS mount path derived from IBM Spectrum Scale. Such a deployment is useful when you are migrating FileNet, running on a traditional Network Attach Storage appliance leveraging its SMB/NFS, to software-defined storage, like IBM Spectrum Scale that also supports SMB/NFS, and a minimum change in the existing architecture is required.

> **Note:** IBM Spectrum Scale 4.2 provides additional protocol access methods in the Standard and Advanced editions. These protocols include NFS, SMB, and Object, which are part of the protocol nodes. In Figure 3 on page 6, the IBM Spectrum Scale protocol nodes, which will host the NFS, SMB, and Object, are not shown separately for the sake of simplicity but are required to be deployed.
>
> For an overview of protocol support with IBM Spectrum Scale 4.2, see this web page:
>
> https://ibm.biz/BdHCU6

IBM Spectrum Scale Object interface: ECM configuration

ECM FileNet Content Manager gives a provision to configure with the back-end OpenStack storage device (Swift). Version 4.1.1 and later of IBM Spectrum Scale supports OpenStack Swift object storage, which supports the REST API interface to consumers and to web applications. Figure 4 represents a basic FileNet Content Manager platform (including Content Platform Engine, Application Engine, Database (DB2)) using the IBM Spectrum Scale Object/REST interface as its OpenStack storage device.
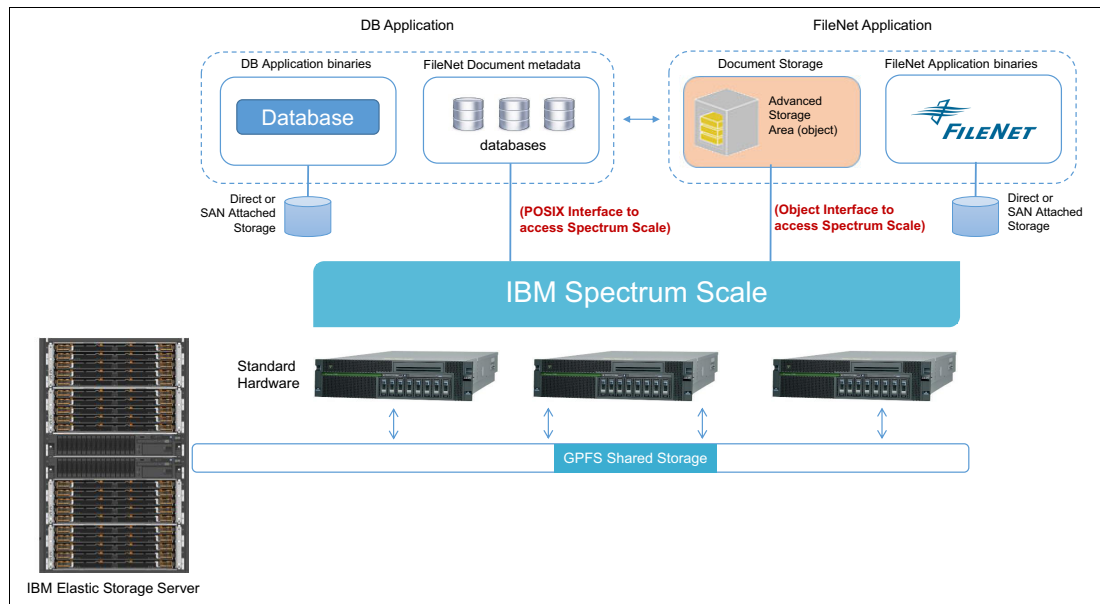

Figure 4   FileNet Content Manager components using IBM Spectrum Scale Object/REST interface

In Figure 4, note that FileNet Content Manager database components are installed, configured using POSIX interface, and the FileNet Advanced Storage Area (named OpenStack storage device) is configured to use the REST interface to store data on IBM Spectrum Scale object store

For more details about OpenStack storage device configuration, see the following web page:

https://ibm.biz/BdXaYi

In some cloud deployment models, applications are required to integrate with cloud object storage. In such deployments, the previously stated setup can be used to enable FileNet to integrate with an object store.

> **Note:** For the deployment, either external Keystone server (from the OpenStack Icehouse or Juno releases, not later than these) must be set up and configured with the IBM Spectrum Scale protocol node as the authentication server. Alternatively, if you prefer to use the Keystone included with IBM Spectrum Scale, you need to have ECM FileNet IBM FileNet Content Platform Engine 5.2.1.3 version.

Deployment suggestion

In all of these deployment topologies, we suggest that the database be hosted on IBM Spectrum Scale POSIX interface for pure performance reasons, although in some deployments, it can also be deployed over NFS.

IBM Spectrum Scale provides additional protocol access methods in the Standard and Advanced editions of the product. These protocols include NFS, SMB, and Object, which are part of the protocol nodes. In the Figure 4 on page 7 deployment diagram, the IBM Spectrum Scale protocol nodes that will host the NFS, SMB, and Object are not shown separately for the sake of simplicity but are required to be deployed.

For an overview of protocol support with IBM Spectrum Scale 4.2, see this web page:

https://ibm.biz/BdHCU6

Value-added features configuration

This section describes the IBM Spectrum Scale features that add value and provide more advantages to ECM solutions by automating and optimizing stored content, and reducing intervention by the storage administrator.

Automated information lifecycle management (ILM)

Information lifecycle management (ILM) can be defined as a set of automated, intelligent tools that use a large set of data that can be classified, managed, organized (placed, migrated) in a cost-efficient way. IBM Spectrum Scale offers a rich set of ILM rules that can be framed on file system attributes (including extended attributes). ILM framework is built in to IBM Spectrum Scale and does not require any third-party integration. ILM features can be used in many variations according to the requirements and can be applied in all topologies that are previously described.

Figure 5 on page 9 demonstrates a basic ILM policy that if the file last access time is younger than a predetermined time then those files are automatically migrated to gold pool solid-state drives (SSD); files that do not fall under this condition are migrated to lower tiers accordingly.
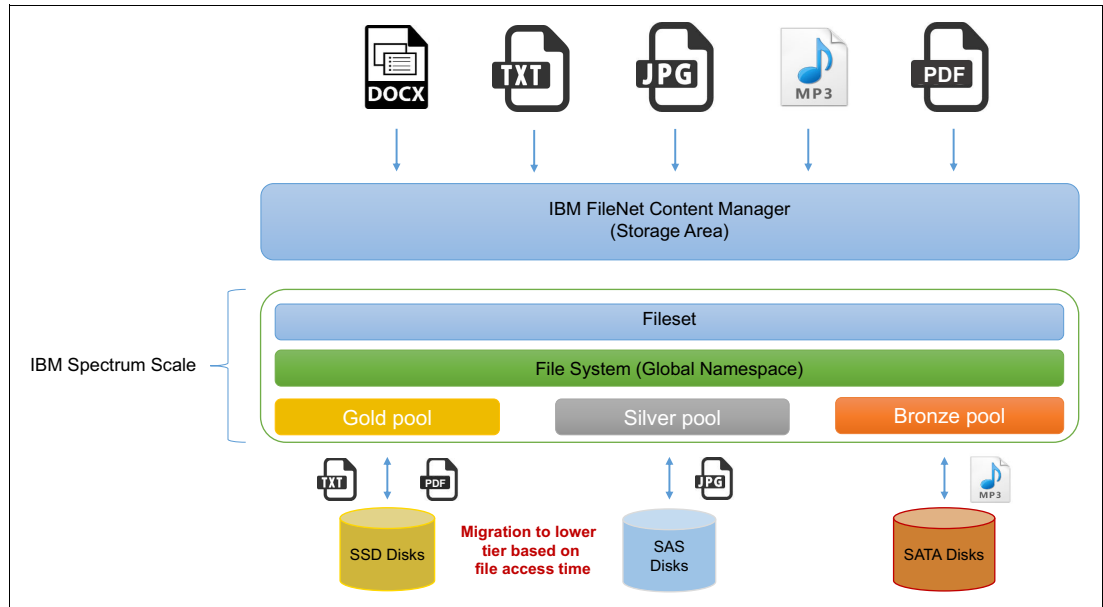
Figure 5   A basic ILM policy usage to migrate files based on access time

For more information about ILM policies, see the following web page:

https://ibm.biz/BdXaYf

With the growth of unstructured data, classifying the data according to importance has become an extremely difficult job for the administrator. But by using the IBM Spectrum Scale ILM policies, data can be efficiently organized across different storage tiers described within a single storage area. However, IBM Spectrum Scale can also help in effective management and cost optimization for structured content using multiple storage areas associated with various storage tiers respectively, as in these examples:

► Storage area1 → storage type: SSD; classification: highly accessed data
► Storage area2 → storage type: SATA; classification: least accessed data

File storage and temperature-based tiering

Figure 6 on page 10 demonstrates a basic ILM policy (file migration rules) that determines whether a file should be automatically migrated to the gold storage pool. If the file's heat is X% compared with other files, then it will be automatically migrated to the gold storage pool (SSD disks). Files not falling under this condition are migrated to lower tiers accordingly. This means that frequently accessed files are migrated to high storage tier whereas least accessed files are migrated to lower storage tier). This is possible by using the file's access temperature, which is an attribute for an IBM Spectrum Scale policy that allows migration of hotter files to higher tiers and colder files to lower tiers.

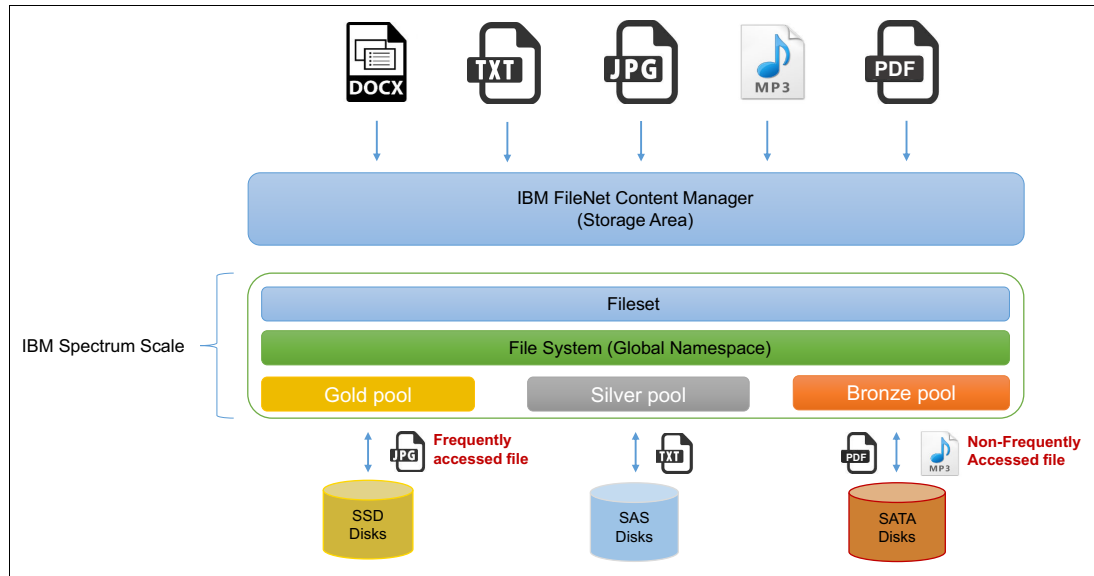For more information, see the following web page:

https://ibm.biz/BdXVZW

Figure 6   A basic ILM policy usage to migrate files based on file heat

> **Suggestion:** IBM Spectrum Scale file management policies based on file access temperature are suggested for optimal use of solid-state drives and ensuring performance for highly accessed ECM hosted data.

Data encryption

From a compliance perspective, encryption is considered one of the most essential features. Data-at-rest encryption has gained more traction today because of the drastic growth of mobile devices, compliance restrictions on offshore archives, and more. The ECM and IBM Spectrum Scale combination offers encryption with high efficiency and easily configurable modes. Using this combination encryption can be performed at two layers:

► Application layer
► Storage layer

At the application layer, FileNet Content Manager Content Platform Engine is configured to encrypt data, whereas IBM Spectrum Scale acts a plain storage back end with no configured encryption, as shown in Figure 7 on page 11. In this mode, when Content Platform Engine retrieves and passes content to a client in response to a client request, the content is automatically decrypted.
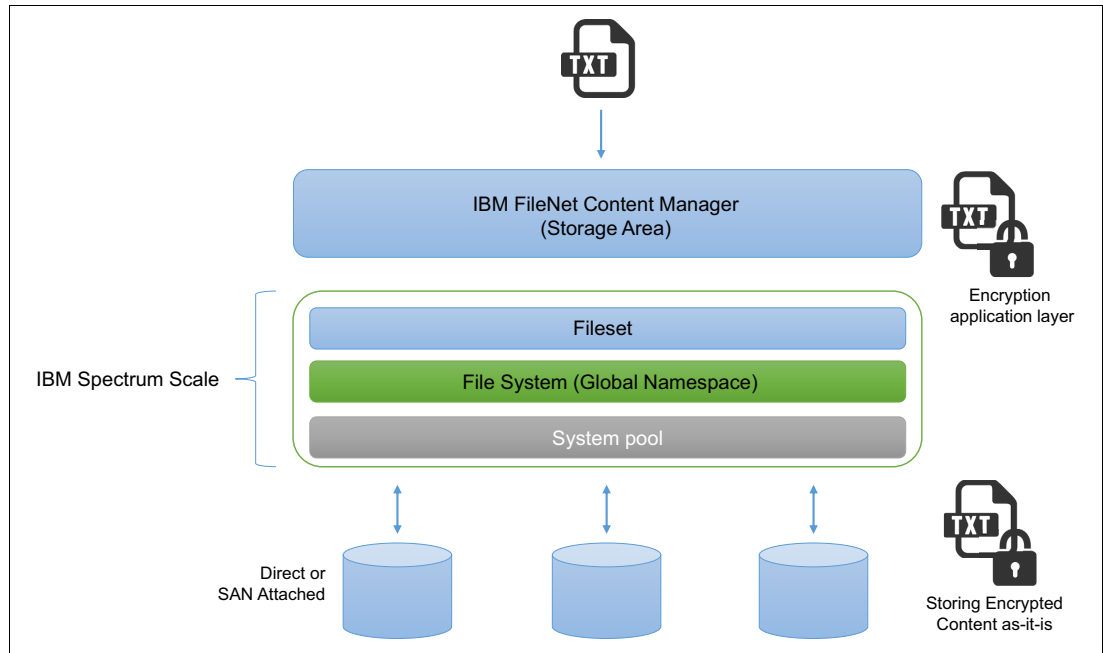
Figure 7   Representation of Encryption at application layer

However, if the file system derived from IBM Spectrum Scale (storage layer) is encrypted and is configured as a storage area, then one does not necessarily need the application-level encryption. In this mode, as shown in Figure 8, FileNet Content Platform Engine does not need to turn on encryption for the storage area because the data is being encrypted at the storage layer.
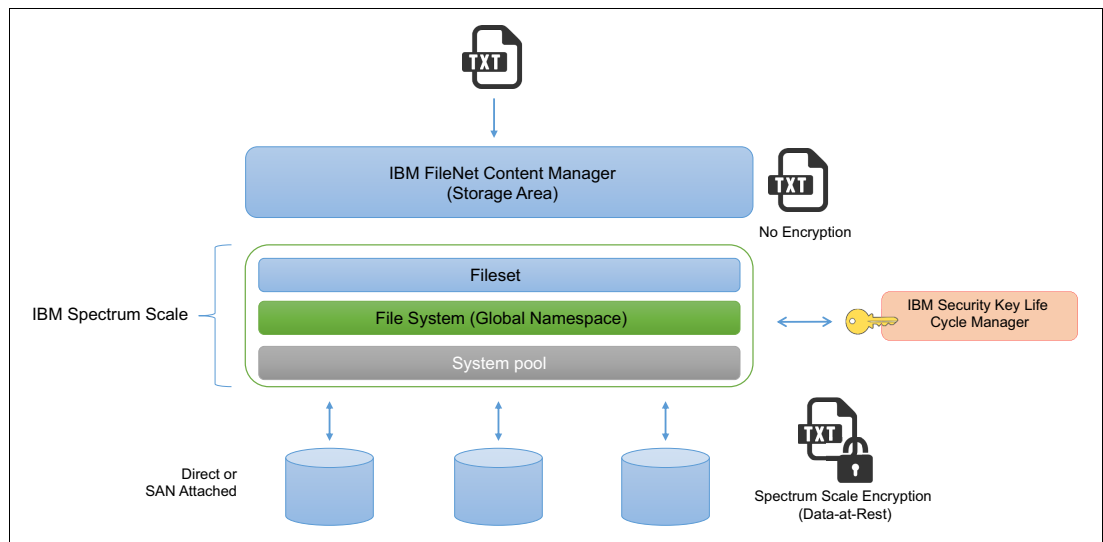
Figure 8   Representation of Encryption at storage layer

In many cases, storage layer encryption results in relatively faster processing of documents due to the encryption job offloaded to the storage controller as opposed to doing encryption at the application layer. But, the key advantage in doing so is to have centralized key management for securing all the data residing on the central back-end IBM Spectrum Scale storage. For example, if in a data lake use case, where IBM Spectrum Scale is used as back-end storage by various applications including that of ECM, one can enable centralized secure data at rest for all the applications by using the IBM Spectrum Scale encryption

feature. This avoids the administration difficulty of managing the encryption configuration at each application level, and more importantly reduces the per-application encryption key management overheads. In addition, encryption with IBM Spectrum Scale also allows for centralized secure deletion of data, required for compliances purposes, which is not supported by application-level encryption.

> **Note:** Secure deletion is a requirement in many compliance-driven workloads where classified data is mist be securely erased so that no data remanence. When ECM data is stored using IBM Spectrum Scale native encryption, data can be securely deleted if required, thus helping host compliance-driven workloads/

See the following web pages for more information:

► IBM Spectrum Scale encryption and secure deletion:

   https://ibm.biz/BdHCUW

► FileNet Content Platform Engine content encryption:

   https://ibm.biz/BdXKnM

Data compression

IBM Spectrum Scale features policy-driven compression to reduce the size of data at rest. Intended primarily for cold data, compression is a background task that occurs after an initial write operation. Data is decompressed on read operation for each access. Using industry-standard compression algorithms, data compression is storage-efficient and CPU-resource-efficient. With IBM Spectrum Scale policies, administrators can compress data only where it will be effective and efficient. Policy-driven compression can be configured to not waste resources that attempt to compress data that is incompressible or decrease the speed of applications by adding overhead to read operations. This way allows ECM FileNet to have its content seamlessly compressed at the back end, thus improving the overall cost effectiveness of the solution.

For more information about IBM Spectrum Scale compression, see the following web page:

https://ibm.biz/BdHC5d

Data integrity using native RAID

IBM Spectrum Scale is also featured with native software RAID, which is available with the IBM Elastic Storage Server. IBM Spectrum Scale native RAID software capability permits to actively manage all RAID functionality formerly accomplished by a hardware disk controller. When ECM is hosted over ESS, the deployment ensures data integrity with enhanced performance and scalability. For more information, see the following resources:

► IBM Elastic Storage Server website:

   http://www.ibm.com/systems/storage/spectrum/ess

► *Introduction Guide to the IBM Elastic Storage Server*, REDP-5253:

   http://www.redbooks.ibm.com/abstracts/redp5253.html

IBM Spectrum Scale performance considerations

The following IBM Spectrum Scale parameters are commonly considered for performance tuning for ECM FileNet workloads:

pagepool

Sufficient pagepool is critical for optimal application performance. IBM Spectrum Scale does not use operating system file cache and will rely on the pagepool for cache, similar to a database bufferpool. Typically, FileNet workloads are categorized under random I/O and therefore, a pagepool of 1 GB or more is known to work well in most cases.

maxFilesToCache

This parameter controls the number of file metadata that can be held in the pagepool. General guidance states that this setting should be large enough to handle the number of concurrently open files in addition to recently used files. Consider full-text indexing of recently ingested content. FileNet Content Manager will need to access the content element after the index request is processed in order to perform extraction. If this content is found in the cache, the overall indexing performance can be improved dramatically. For scenarios where recently ingested content is not required for immediate retrieval, the default value should suffice. In IBM Spectrum Scale 4.1 and later, `maxFilesToCache` defaults to 4000.

maxStatCache

This parameter sets aside pageable memory to cache attributes of files that are not currently in the regular file cache. In IBM Spectrum Scale 4.1 and later, `maxStatCache` defaults to 1000. If `maxFilesToCache` is highly increased, then the preference is for `maxStatCache` to be manually set to a value less than `4 * maxFilesToCache`, although the default value is sufficient in most cases. FileNet CM workloads that are heavily oriented toward content retrieval, might see improved performance by increasing this value. On Linux platforms, the way Linux handles inodes makes `maxStatCache` generally ineffective. So, on Linux systems, keep `maxStatCache` at the default of 1000 and modify `maxFilesToCache` as needed.

maxMBps

The preference is to set this parameter to twice the throughput required by the system. In IBM Spectrum Scale v3.5 and later, the default is 2048 MB per second, which should be sufficient for most applications.

blocksize

The default file system block size (256K) should be sufficient for most applications, although you may consider a larger block size if large content is being stored on the file system. For maximum performance, always consider the IBM Spectrum Scale block size when defining the RAID stripe size for the underlying LUN or LUNs. The IBM Spectrum Scale block size should match, or be a multiple of, the RAID stripe size.

Separate disks for data and meta-data

With IBM Spectrum Scale, you can define a set of disks to hold only data, only file system metadata or both data and metadata. For general purposes, separating metadata from data is not necessary. If your FileNet Content Manager deployment frequently performs file system metadata operations, application performance might be improved by dedicating multiple disks for only file system metadata. In such cases, the preference is to have file system metadata on solid-state drives.

For details and related information about IBM Spectrum Scale tuning parameters and ECM recommendations, see the FileNet Content Manager and GPFS web page:

http://www.ibm.com/support/docview.wss?uid=swg27042971&aid=1

Case study

This section describes a case study where the IBM ECM product suite is deployed along with IBM Spectrum Scale. This is a medium to large deployment for a multi-national customer supporting hundreds of terabytes of storage.

The problem

A large telecommunications company had put forth the following high-level requirements.

The telecom service provider company has a customer base throughout the country, spread across a total of 23 US states. The customer base at that time was hovering at approximately 85 - 90 million and grew to 160 million. Each of these customers submits a set of documents when registering for the services provided by the telecom service provider. The documents must be verified, cataloged, and maintained. The government authority needed and continues to need a mechanism to access and audit this data on occasion; the query and access to this data can go as far back in time as 15 years. This means, the data must be made available on demand and within a stipulated period of time. The data is required to be stored separately, per state of the country, in an ever-increasing scalable platform. The system must also be designed to handle the load of daily ingestion of customer data, amounting to approximately 50 - 70 GB per day. Additionally, the client also has a backlog of approximately 80 TB or more of customer data to be loaded in the system.

The solution

IBM FileNet met the customer requirements because of the following benefits:

► Consists of components such as Content Platform Engine, which is a FileNet Content Manager component that is designed to handle the heavy demands of a large enterprise.
► Can manage enterprise-wide workflow objects, custom objects, and documents by offering powerful and easy-to-use administration tools.
► The tools help the administrator easily create and manage the classes, properties, storage, and metadata that form the foundation of an ECM system.

The scalable robust ECM platform was chosen because it can o store, search, and retrieve this data.

FileNet stores its content in two formats: metadata and document content. In this deployment, customer identification data was stored as metadata in a relational database engine; the other customer data (such as scanned customer documents) were stored in file systems in an encrypted format. This segregation of metadata and content fits perfectly to solve this problem; it can scale efficiently with easy lookup of the data.

Specifically for handling the large load of millions of files, IBM Spectrum Scale was chosen to work with FileNet. IBM Spectrum Scale along with FileNet provided the much required scalable, enterprise class document management solution, with the ability to easily extend to petabytes, meet the on-demand access to consumer data within a stipulated period of time, and provide the required encryption to customer data.

For this deployment, which is now over 500 TB and growing, the enterprise-level content management solution based on FileNet and IBM Spectrum Scale, proved to be a winning combination that successfully met all the customer requirements. The solution continues to be stable, scalable and cost-efficient.

Figure 9 shows the high-level logical architecture of the solution. A single IBM Spectrum Scale cluster was formed and a dedicated file system for each US state was assigned. Assigning a dedicated file system for each state allowed for having a storage quota for each state and also having scheduled backup for each state. Another way of achieving this requirement is to create one single IBM Spectrum Scale file system and have each state assigned to an independent IBM Spectrum Scale fileset.
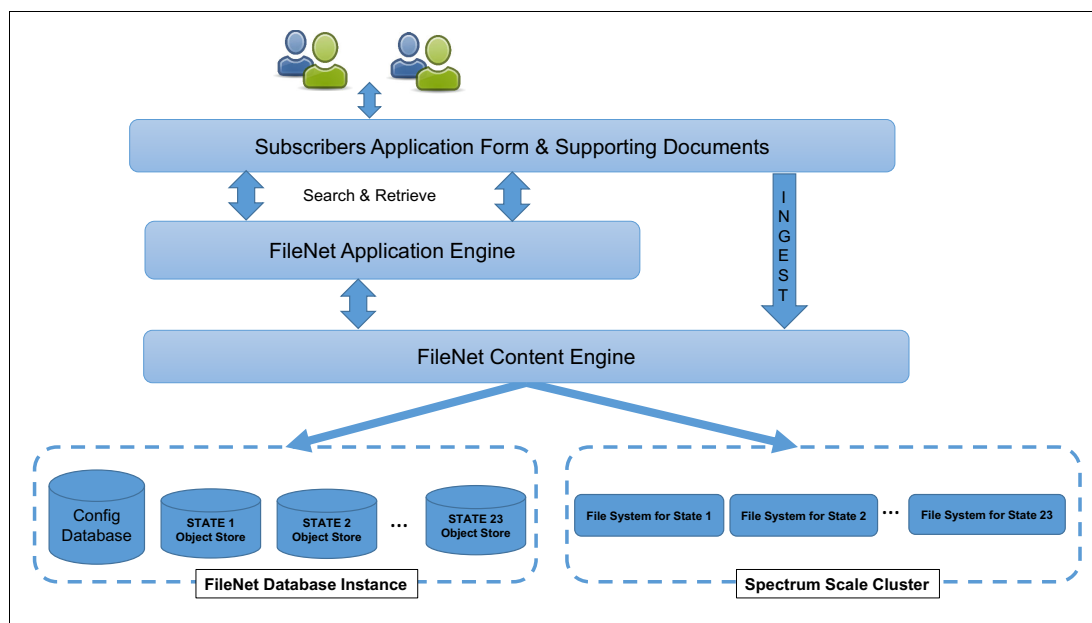

Figure 9   Representation of high-level logical architecture at the customer location

Sample deployment details for setups

The deployment topology chosen was "IBM Spectrum Scale POSIX interface: ECM configuration" on page 5. For this case study, the shared storage was addressed by six IBM Spectrum Scale nodes based with SAN-based block storage. Recently released IBM Elastic Storage Server is a more cost-efficient back-end storage because of its declustered RAID functionality, and can be used in such deployments.

Detailed configuration steps for IBM Spectrum Scale cluster and ECM are not in the scope of this paper. See their respective product documentation. However, for reference and planning, following are some sample configuration listings associated with such deployments.

IBM Spectrum Scale cluster information

Example 1 displays the current configuration information for the use case GPFS cluster using the `mmlscluster` command.

Example 1   Current configuration information

```
Output of mmlscluster
GPFS cluster information
========================
  GPFS cluster name:         FileNet_ProdN.filenet_gpfs3
  GPFS cluster id:           12577970869153660504
  GPFS UID domain:           FileNet_ProdN.filenet_gpfs3
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
```

```
GPFS cluster configuration servers:
-----------------------------------
  Primary server:    filenet_gpfs3
  Secondary server:  filenet_gpfs4

 Node  Daemon node name  IP address      Admin node name  Designation
--------------------------------------------------------------------------
    1    filenet_gpfs1     192.168.107.27  filenet_gpfs1    quorum-manager
    2    filenet_gpfs2     192.168.107.28  filenet_gpfs2    quorum-manager
    3    filenet_gpfs3     192.168.107.29  filenet_gpfs3    quorum-manager
    4    filenet_gpfs4     192.168.107.30  filenet_gpfs4    quorum-manager
    5    filenet_gpfs5     192.168.107.49  filenet_gpfs5
    6    filenet_gpfs6     192.168.107.50  filenet_gpfs6
```

Define tiebreaker disks for a two-node cluster

Tiebreaker disks are suggested when you have a two-node cluster or you have a cluster where all nodes are SAN-attached to a common set of logical unit numbers (LUNs) and you want to continue to serve data with a single surviving node. Tiebreaker disks are not special NSDs; you can use any NSD, which is connected to all quorum nodes, as a tiebreaker disk. Although you can have one, two, or three tiebreaker disks, you should use an odd number of tiebreaker disks. The `mmchconfig` command with the `tiebreakerDisks` option is used to set tiebreaker disks (Example 2). Be sure to stop the IBM Spectrum Scale cluster while tiebreaker disks are set. Because the deployment grew from a two-node cluster, tiebreaker disks were used.

Example 2   Using the mmchconfig command to set the tiebreaker disks

```
#  mmumount    all    -a
#  mmshutdown   -a
#  mmchconfig   tiebreakerDisks="aecnsd1"
#  mmstartup    all
#  mmstartup    -a
```

Set appropriate pagepool

The `pagepool` parameter of the `mmchconfig` command determines the size of the IBM Spectrum Scale (GPFS) file data block cache. Unlike local file systems that use the operating system page cache to cache file data, Spectrum Scale allocates its own cache called the *pagepool*. The IBM Spectrum Scale pagepool is used to cache user file data and file system metadata.

The pagepool size used for such deployments can be 4096m, which can be changed by using the `mmchconfig` command:

`# mmchconfig pagepool=4096m`

IBM Spectrum Scale file system attributes

The `mmlsfs` command displays the file system attributes of the file system configured for such a deployment (Example 3).

Example 3   File system attributes displayed using the mmlsfs command

```
File system attributes for fs1:
===================================

flag                value                   description
------------------- ----------------------- -------------------------------------
 -f                 8192                    Minimum fragment size in bytes
 -i                 512                     Inode size in bytes
 -I                 16384                   Indirect block size in bytes
 -m                 1                       Default number of metadata replicas
 -M                 2                       Maximum number of metadata replicas
 -r                 1                       Default number of data replicas
 -R                 2                       Maximum number of data replicas
 -j                 cluster                 Block allocation type
 -D                 nfs4                    File locking semantics in effect
 -k                 all                     ACL semantics in effect
 -n                 32                      Estimated number of nodes that will
mount file system
 -B                 262144                  Block size
 -Q                 none                    Quotas enforced
                    none                    Default quotas enabled
 --filesetdf        no                      Fileset df enabled?
 -V                 10.01 (3.2.1.5)         File system version
 --create-time      Sun Jul 11 13:06:51 2010 File system creation time
 -u                 yes                     Support for large LUNs?
 -z                 no                      Is DMAPI enabled?
 -L                 4194304                 Logfile size
 -E                 yes                     Exact mtime mount option
 -S                 no                      Suppress atime mount option
 -K                 whenpossible            Strict replica allocation option
 --fastea           no                      Fast external attributes enabled?
 --inode-limit      15001600                Maximum number of inodes
 -P                 system                  Disk storage pools in file system
 -d                 fs01_nsdDS2;fs01_nsdDS1 Disks in file system
 --perfileset-quota no                      Per-fileset quota enforcement
 -A                 yes                     Automatic mount option
 -o                 none                    Additional mount options
 -T                 /fs1                    Default mount point
 --mount-priority   0                       Mount priority
```

FileStore configuration from FileNet

Figure 10 shows the configuration from the FileNet Administrative Console for Content Engine (ACCE) from the sample setup. It shows the list of various file storage areas defined within the system for the states. This view shows the number of files and the space occupied by these files on the file storage area.

		Display Name ▲	Type	Status	File Count	File Bytes	Fixed Content Device	Site
☐		Default Database Storage Area	Database Storage Area	Open	0	0 KB		Initial S
☐		VFNAPOS1	File Storage Area	Closed	292,358	41278 MB		Initial S
☐		VFNAPOS10	File Storage Area	Closed	1,060,218	125703 MB		Initial S
☐		VFNAPOS100	File Storage Area	Closed	123,041	28494 MB		Initial S
☐		VFNAPOS101	File Storage Area	Closed	407,808	115017 MB		Initial S
☐		VFNAPOS102	File Storage Area	Closed	80,998	22869 MB		Initial S
☐		VFNAPOS104	File Storage Area	Closed	47,860	13369 MB		Initial S
☐		VFNAPOS105	File Storage Area	Closed	33,452	9241 MB		Initial S
☐		VFNAPOS106	File Storage Area	Closed	246,488	75504 MB		Initial S
☐		VFNAPOS107	File Storage Area	Closed	18,190	5693 MB		Initial S

New   Refresh   Batch Operations

Figure 10   Sample configuration from FileNet Administrative Console for Content Engine (ACCE)

Figure 11 shows the mapping between FileNet file storage areas and mount point or file system from the operating system (OS). Display name is what is seen or displayed in the FileNet environment. Root directory path is the entry from the file system (or mount point) that must be referred to. After the file system in this way is mapped to FileNet, it is available for storing content. In the deployment, there is a file system mount point per state.

Save   Refresh   Actions ▾   Close

File Storage Area: VFNAPOS1

General   Properties   Configuration   Statistics   Storage Policies

Define the storage area by providing a name and description and site information.

* Display name: ⑦ VFNAPOS1

Description: ⑦ VFNAPOS1

Id: ⑦ {DB6215F3-FB6C-44D1-BE54-47C6D64F9078}

* Root directory path: ⑦ /fs7/vfnapos/vfnap1

Site: ⑦ Initial Site

Object store: ⑦ VFNAPOS

Figure 11   Mapping between FileNet file storage areas and mount point or file system from the OS

Summary

IBM FileNet Content Manager is a content, security, and storage management engine with ready-to-use workflow and process capabilities. It helps meet the growing challenge of managing enterprise content by providing a single repository for enterprise content. It streamlines content management and delivery with document management services, and delivers content in motion for better business value and lower costs.

IBM Spectrum Scale is a proven, scalable, high-performance data and file management solution that is being used extensively across multiple industries worldwide. IBM Spectrum Scale provides simplified data management and integrated information lifecycle tools capable of managing petabytes of data and billions of files in order to manage the growing cost of dealing with ever-increasing amounts of data across your organization.

Combining FileNet Content Manger with IBM Spectrum Scale can maximize your document storage capabilities while minimizing IT costs. This paper illustrates the various deployment topologies to consider when you plan a FileNet Content Manager installation. In addition to storing critical business content, IBM Spectrum Scale provides multiple value-added features that help automate, encrypt, and optimize stored content in addition to lowering administration overhead.

Other resources for more information

See the following resources for more information:

► IBM Spectrum Scale resources

http://www.ibm.com/systems/storage/spectrum/scale/resources.html

► IBM Spectrum Scale in the IBM Knowledge Center

http://www.ibm.com/support/knowledgecenter/SSFKCN/gpfs_welcome.html

► IBM Spectrum Scale Overview and Frequently Asked Questions (FAQ)

http://ibm.co/1IKO6PN

► IBM Spectrum Scale wiki

https://ibm.biz/BdXVxv

► IBM ECM resources

https://ibm.biz/BdXyh8

► IBM Analytics

http://www.ibm.com/analytics/us/en/

► *IBM FileNet P8 Platform and Architecture*, SG24-7667

http://www.redbooks.ibm.com/abstracts/sg247667.html?Open

► *IBM High Availability Solution for IBM FileNet P8 Systems*, SG24-7700

http://www.redbooks.ibm.com/abstracts/sg247700.html?Open

Authors

This paper was produced by a team of specialists from around the world working with the International Technical Support Organization, Tucson Center.

Sandeep R. Patil is a Senior Technical Staff Member and working as a Storage Architect with IBM System Labs. He has over 15 years of extensive product architecture and design experience. Sandeep is an IBM Master Inventor, an IBM developerWorks® Master Author, and also a member of the IBM Academy of Technology. Sandeep holds an Bachelor of Engineering (Computer Science) degree from the University of Pune, India.

Atul Gore is an Information Technology Professional with IBM Enterprise Content Management Lab Services and has a total of 22 years of IT experience. He is IBM FileNet P8 certified and is an expert in IBM Content Manager Enterprise Edition who has successfully designed, installed, configured, maintained, and supported various IBM Enterprise Content Management customer solutions worldwide. He holds a Masters degree in Computer Science from the University of Pune, India.

Sasikanth Eda is a Software Engineer with the IBM Spectrum Scale development team. He works for the integration of OpenStack with IBM Spectrum Scale, focusing on Swift object, Cinder block, and Manila file storage components of OpenStack. Sasi holds a Masters degree in Microelectronics from I2IT, Pune, India.

Michael Bordash is a Software Engineer with the IBM Analytics Enterprise Content Management development team in Costa Mesa, California, US. He is a team lead within ECM system test, responsible for the performance and scalability of IBM FileNet Content Manger. He joined IBM in 2005 and has worked on the IBM FileNet suite of products for the past seven years. Michael holds a B.Sc. in Computer Science from Appalachian State University.

Sanjay K. Sudam is a Solution Architect with the IBM Storage ISV Technical Enablement team. He is responsible for the technical enablement of the various ISV partner solutions with the IBM Storage portfolio including flash, block, file, and object storage. Sanjay holds a degree in Electronics and Communication Engineering from the Nagarjuna University, India.

Sathish Subramanyam is a Senior ECM Architect in the IBM India Software Labs. He has extensive experience of implementing a wide range of ECM solutions. He currently works with the Content Integrator and Image Services Adapter products within the FileNet ECM Suite. He holds a Masters Degree in Software Systems and recently completed an Executive Program in General Management.

Thanks to the following people for their contributions to this project:

Larry Coyne
International Technical Support Organization, Tucson Center

Duane Baldwin, Bob Kreuch, David Lebutsch
IBM Analytics, Platform

Michael Daubman, Pallavi Galgali, Wei G. Gong, Nils Haustein,
Vincent Hsu, Tomer Perry, Arati Ranade, Yuri L Volobuev
IBM Systems

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- ► Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- ► Follow us on Twitter:

  http://twitter.com/ibmredbooks

- ► Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- ► Explore new IBM Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- ► Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

23

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®	IBM®	Redpaper™
DB2®	IBM Elastic Storage™	Redbooks (logo) ®
developerWorks®	IBM Spectrum™	z Systems™
FileNet®	IBM Spectrum Scale™	
GPFS™	Redbooks®	

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Printed in U.S.A.

Get connected

ibm.com/redbooks