Enabling easy movement of medical imaging data across hybrid cloud tiers

Use policy-based data movement across tiers to place inactive data into IBM Cloud Object Storage

Overview

Challenge

- Delivering the best quality of patient centric healthcare
- Providing healthcare at the lowest cost
- Ageing populations
- Encountering the burden of chronic diseases
- Increasing costs imposed on current human, clinical, and social resources

Solution

IBM combines the healthcare VNA applications with the IBM Spectrum Scale software, IBM Transparent Cloud Tiering, and IBM Cloud Object Storage to offer:

- A standards-based, secure, and robust architecture
- A high performance active data storage
- A simple policy-based data movement across hybrid cloud tiers
- IBM Cloud Object Storage for inactive data

Healthcare executives are pursuing a journey of transformation from prescriptive clinical advice to evidence-based medicine by focusing on patient wellness and outcomes, and minimizing the volume of services consumed. They are also adopting technology, process, and workflow improvement strategies to enable secure access, exchange, and analysis of patient information.

This paper provides recommendations and best practices to help ensure an efficient installation of the solution to use policies to move any Vendor Neutral Archive (VNA) or medical imaging data from Spectrum Scale to IBM Cloud Object Storage, with acceptable performance.

Introduction

Healthcare providers are using cloud, Data Lakes, analytics, and cognitive solutions to deliver best-in-class patient care. They are currently making use of block, file, and cloud to store millions of data files originating from thousands of medical imaging procedures, laboratory tests, and clinical procedures. Patient studies used to be files that are used once then filed away forever. This is unstructured data, and therefore, it cannot be easily seen, and it results in the creation of inefficient data islands within the departmental silos of a healthcare organization.

Many healthcare providers are transitioning to software-defined frameworks such as IBM® Spectrum Scale™ and IBM Watson® cognitive solutions to deliver significantly higher levels of agility, security, cost-effectiveness, flexibility, and efficiency as the key deterministic priorities for their IT infrastructures. They are eliminating departmental silos by implementing standards-based, highly

interoperable solutions, and to help them deliver the highest quality of care to their patients at a lower cost.

Table 1 highlights the journey, from departmental silos to cognitive health.

Environment variables	Traditional departmental deployments (unstructured data)	Enterprise data lakes delivering analytic insights (by IBM Spectrum Scale)	Cognitive information architecture (by IBM Watson)
Clinical example	**Clinical advice**: Tumor mass suspected in the X-Ray	**Clinical informatics guidance**: How many tumors/cancers of certain kind and size were treated in last six months?	**Evidence-based medicine**: Is the cancer shrinking faster than other patient histories? Tailor clinical services if shrinkage is (or is not) occurring at the expected rate.
Architectural topology	**Traditional architecture**: Enterprise imaging archives to store unstructured data.	**Software-defined architecture**: Data Lakes supporting analytic applications	**Cognitive solutions architecture**: Leverage SDI as a foundation with flexible services – based on natural language, algorithms, machine learning, and analytical reasoning capabilities based on learned experience
Results	**Busy work**: Unpredictable outcomes	**Improved process and workflow efficiencies**: Higher quality of care, and minimized capital expenditures	**Healthier, happier populations**: Longer, more productive lives

Table 1: The journey to cognitive health services

Solution configurations, tests, and results

The storage solution

The IBM storage solution consists of IBM Spectrum Scale running on IBM Elastic Storage Server (ESS) hardware for high performance, IBM Transparent Cloud Tiering running on x86 Red Hat Enterprise Linux (RHEL) servers for the cloud transport layer, and IBM Cloud Object Storage to store the GE Clinical Archive inactive data.

Configurations and tests

Through this configuration and test (refer to Figure 1), IBM Spectrum Scale was installed on an IBM Elastic Storage Server. A file share was configured and enabled as a library for the VNA software application that was configured as a virtual machine (VM) on an ESXi server. The VNA application used in this effort was GE Healthcare Centricity Enterprise Archive software. It serves as a foundation for Centricity Clinical Archive VNA software and Centricity Universal Viewer software.

The VMs were configured with Microsoft® Windows® 2012 software pre-packaged in self-installable, ready to deploy, VMWare OVF template formats, provided by GE Healthcare Corporation.

Architecture

Software
- IBM Spectrum Scale 4.2.1
- IBM Elastic Storage Server 4.5.0
- IBM Transparent Cloud Tiering 1.1.1
- IBM Cloud Object Storage 3.8.0.134

Hardware
- IBM Elastic Storage Server GL6
- IBM Cloud Object Storage appliance
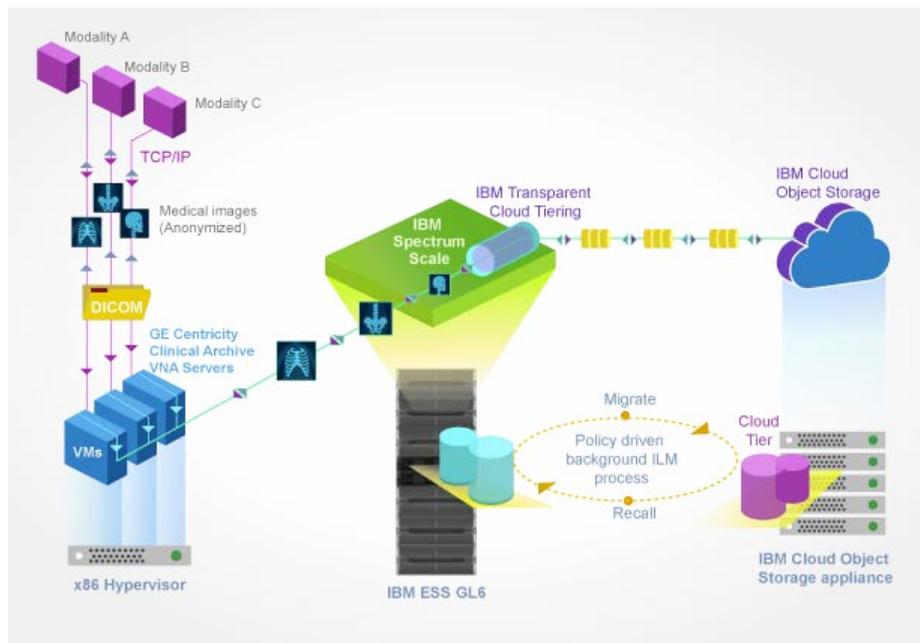- Lenovo System x3850 X5 server

---

Figure 1: Policy-based data flow across IBM Spectrum Scale, IBM Transparent Cloud Tiering, and IBM Cloud Object Storage

Test observation and results

- The GE Healthcare Centricity Enterprise Archive software was successfully installed, configured, and tested with IBM Spectrum Scale software, using the Server Message Block (SMB) protocol for the file share.
- A policy-driven background information lifecycle management (ILM) process moved the image files to and from Spectrum Scale through IBM Transparent Cloud Tiering to IBM Cloud Object Storage.
- When GE Centricity Clinical Archive requests an image from Spectrum Scale that has been migrated to IBM Cloud Object Store, Spectrum Scale automatically retrieves it from the IBM Cloud Object Storage through Transparent Cloud Tiering.
- Functionality, reliability, and availability tests were run in detail, and all features were tested successfully.

File systems layout recommendations

A distinction is made between short-term online storage and long-term archive storage as follows:

- **Short-term storage (STS):** New images are sent from the modality to the PACS or VNA server, and then stored in the Spectrum Scale 4.2.2 file system, directly.
- The PACS or VNA server communicates with the Spectrum Scale 4.2.2 file system, using the SMB protocol, and then writes the images into the Spectrum Scale filesystem, that acts as the STS.
- **Long-term storage (LTS):** Images are archived from the short-term storage to the long-term storage based on information lifecycle management (ILM) policies. It should be noted, that the movement of the DICOM file is transparent to the PACS or the VNA application. The DICOM file is moved from STS to LTS archive; however, the policy leaves a small stub on the STS. A stub is a tiny amount of file data that points to the relevant address, from where to fetch the file back. When a Universal Viewer or a Clinical Console accesses the STS resident stub, it will fetch the file back from LTS, and furnish it to the Universal Viewer/Clinical Console. The long-term storage provides a very large storage capacity. Images can be archived there for many years.
- A policy driven background ILM process moves the image files to and from Spectrum Scale through IBM Transparent Cloud Tiering to IBM Cloud Object Storage.
- If PACS or VNA requests an image from Spectrum Scale, that is currently *not resident* on STS, then Spectrum Scale automatically retrieves it from IBM Cloud Object Storage through IBM Transparent Cloud Tiering.

Guidance and recommendations for policy-based data movement

The backbone of this solution are the Spectrum Scale policy files that can be easily applied to move, manage, and migrate data

A *policy* is a set of rules that describes the life cycle of user data based on the attributes of files.

The operations to be performed are defined in the rules. The operations can be:

- File placement
- Migration
- Listing
- Compression
- Encryption
- Deletion

The Spectrum Scale policy engine identifies files to be archived based on its attributes and the criteria defined for the selection in the rules and then performs the operation defined in the rule for that specific criteria. In this paper, migration and recall policies are discussed, that are unique to this solution.

IBM Transparent Cloud Tiering and policy
IBM Transparent Cloud Tiering takes advantage of the Spectrum Scale policy engine to enable data movement from Spectrum Scale to on-premises or public cloud and back to Spectrum Scale.

Transparent Cloud Tiering has two ways to migrate or recall the files to and from Spectrum Scale file system and cloud storage.

Transparent or automatic migration
Rules are defined in an installed policy to transparently migrate data to the cloud based on the criteria defined in the attributes of the files. The default policy can be defined using the `mmchpolicy` command. The Spectrum Scale policy engine migrates the file to cloud storage keeping only the file stub on the Spectrum Scale file system. A stub is a tiny amount of file data that points to the relevant address, from where to fetch the file back. When a Universal Viewer or a Clinical Console accesses the STS resident stub, it will fetch the file back from LTS, and furnish it to the Universal Viewer/Clinical Console.

Manual migration

You can use the `mmcloudgateway` or the `mmapplypolicy` command, to migrate files from one tier of storage to another.

Transparent or automatic recall

Accessing an archived file triggers transparent or automatic recall of data from cloud storage. The file data that have been migrated to the cloud storage are copied back to Spectrum Scale keeping a copy on cloud storage.

Manual recall

Using the `mmcloudgateway` or the `mmapplypolicy` command, a file can be recalled from the cloud storage to the Spectrum Scale file system.

It is our observation, from our tests, that Migration or recall using the `mmapplypolicy` command gives better performance than directly running the `mmcloudgateway` command.

You can find more information on how to write and use Spectrum Scale policies at:
ibm.com/support/knowledgecenter/STXKQY_4.2.0/com.ibm.spectrum.scale.
v4r2.adv.doc/bl1adv_policies.htm

Restore

The files that are corrupted or deleted locally, can be restored from the LTS archive resident on IBM Cloud Object Storage back to Spectrum Scale using the `mmcloudgateway` command with the `restore` option. You can restore the file to the original location or to a new location, on the STS, Spectrum Scale filesystem.

Solution benefits – summary

The software-defined solution (IBM Spectrum Scale, IBM Transparent Cloud Tiering, and IBM Cloud Object Storage) offers the following significant benefits for healthcare clients running VNA application in their environments:

- Unifies departmental archives of unstructured data, including independent data stores, into a common, single, highly available and secure global namespace

- Manages patient information at a virtually limitless scale and mines that data for new insights to improve patient outcomes

- Uses software intelligence to automatically migrate medical data to the optimum storage tier based on performance and cost objectives

- Lowers the cost of information management significantly with a shared storage infrastructure

- Enables collaboration of various clinical specialties, across the enterprise with a shared VNA

- Manages medical image data growth on a limited budget

Acknowledgments

Special thanks to the GE Healthcare product management and engineering teams in the United States and Europe for loaning the software and licenses that enabled the IBM engineering teams to successfully create an operational test environment and run tests to document real-life results.

Many thanks to the IBM Offering Management, IBM Spectrum Scale Development, IBM Transparent Cloud Tiering Development, IBM Cloud Object Storage Development, IBM Client Executives, IBM Systems members, and other associates worldwide, who contributed with their recommendations and excellent guidance, during the test run and review process and enabled successful completion and validation of this qualification project.

Resources

The following websites provide useful references to supplement the information contained in this paper:

- IBM Systems on PartnerWorld
  ibm.com/partnerworld/systems
- IBM Redbooks
  ibm.com/redbooks
- IBM Publications Center
  www.elink.ibmlink.ibm.com/public/applications/publications/cgibin/pbi.cgi?
  CTY=US

- IBM Spectrum Scale Knowledge Center
  http://www.ibm.com/support/knowledgecenter/STXKQY
- IBM Spectrum Scale product documentation and related publications
  www.ibm.com/support/knowledgecenter/STXKQY_4.2.0/ibmspectrumscale42_welcome.html
- Elastic Storage Server (ESS), see IBM Knowledge Center
  ibm.com/support/knowledgecenter/SSYSP8/sts_welcome.html
- IBM General Parallel File System (IBM GPFS™)
  ibm.com/support/knowledgecenter/SSFKCN/gpfs_welcome.html
- IBM POWER8 processor-based servers
  ibm.com/support/knowledgecenter/POWER8/p8hdx/POWER8welcome.htm
- Configuring Protocols Quick Overview for Spectrum Scale
  ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/General%20Parallel%20File%20System%20%28GPFS%29/page/Protocols%20Quick%20Overview%20for%20IBM%20Spectrum%20Scale
- Extreme cluster/cloud administration toolkit (xCAT), go to the xCAT website
  http://sourceforge.net/p/xcat/wiki/Main_Page/
- System Storage DCS3700 Quick Start Guide, GA32-0960-03
  ibm.com/support/docview.wss?uid=ssg1S7005178&aid=1
- IBM System Storage DCS3700 Storage Subsystem and DCS3700 Storage Subsystem with Performance Module Controllers: Installation, User's, and Maintenance Guide, GA32-0959-07
  ibm.com/support/docview.wss?uid=ssg1S7004920

About the authors

Prashant Avashia is a senior architect for storage and software-defined infrastructure in IBM Systems Group. He has architected and delivered cloud, analytics, and software-defined solutions for global healthcare clients. You can reach Prashant at pavashia@us.ibm.com

Sandeep Zende is an architect for healthcare solutions in IBM Systems Group. He has architected and delivered storage migrations, disaster recovery solutions, and software-defined solutions for several clients, including healthcare clients. You can reach Sandeep at sandeep.zende@in.ibm.com.

Barry VanSloten is an ex-IBMer. While working for IBM, he was an architect in the IBM Systems Group. He had designed, implemented, and consulted on storage and server architectures, implementations, and migrations for clients in various industries, including healthcare.