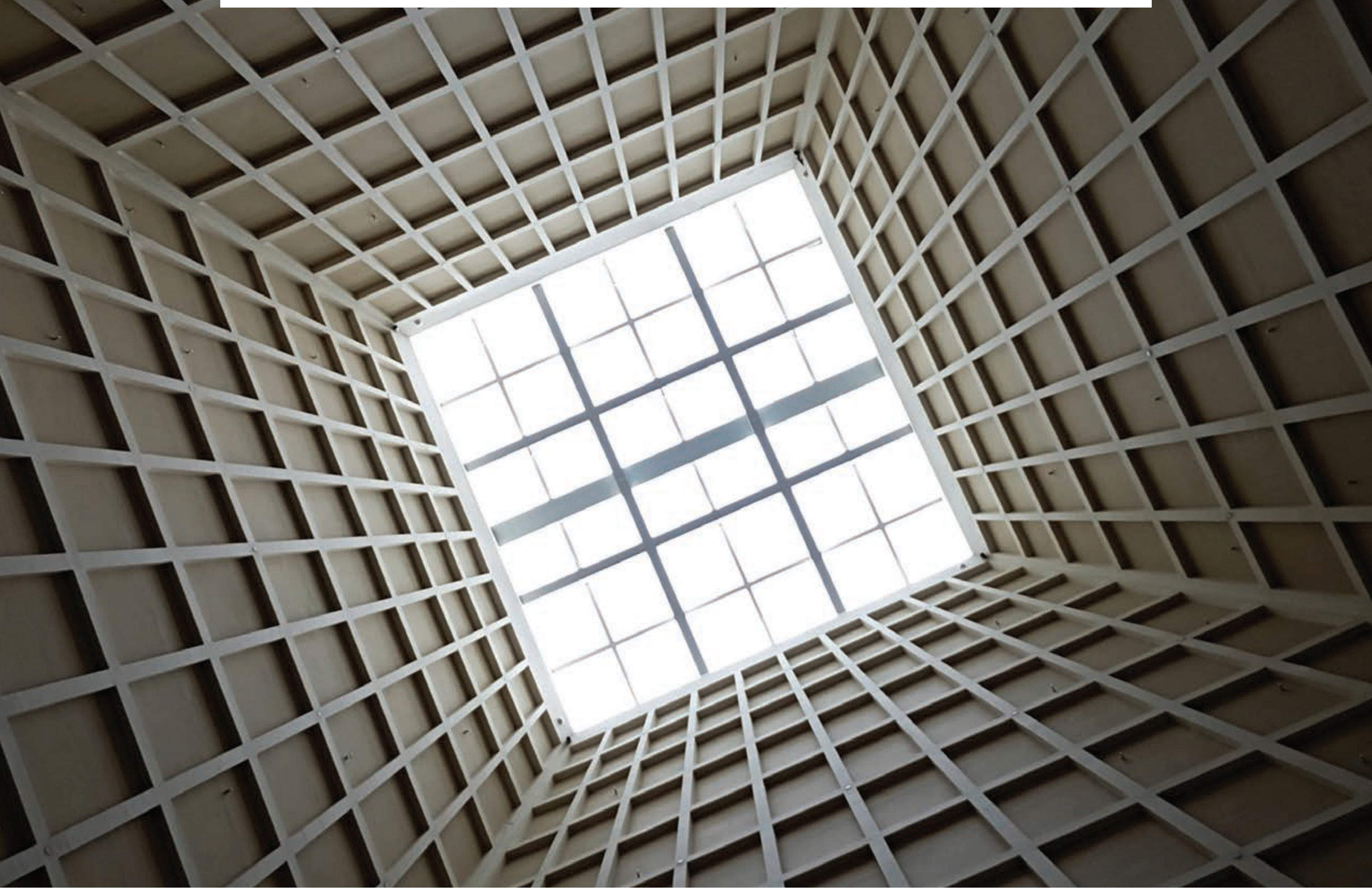
WHITE PAPER

Maximizing the DICOM Standard for Digital Pathology Routing, Ingesting, and Storing



A Dicom Systems and Pure Storage White Paper

Publication: September 2024

Background Summary

Adopting the Digital Imaging and Communications in Medicine (DICOM) standard in pathology has typically been driven by the crucial need for interoperability among various slide scanners and IT systems. The DICOM standard offers superior performance in displaying, viewing, and storing images, achieved by dividing large images into smaller tiles stored in two-dimensional arrays. This method allows for faster image retrieval by enabling quick access to specific sub-regions without loading extensive data. Digital Whole Slide Images (WSI) management has many pathology organizations seeking scalable, simple, and cost-optimized solutions experienced in DICOM and non-DICOM.

Processing digitized pathology slides into DICOM format and routing the images across the organization significantly affects the network and the DICOM application handling that traffic. Bottlenecks may arise from processing, routing, transformation, and impact of artificial intelligence (AI) algorithms used in imaging workflows. Multiple factors contribute to these challenges, such as complex imaging workflows, disparate systems operating in varied environments, outdated enterprise imaging platforms, or inadequate data storage for high-volume imaging.

With the average DICOM Pathology file size being several factors larger than traditional Radiology and Cardiology DICOM data sets, storage consumption becomes a critical problem to solve. According to ***J Pathol Inform,*** the DICOM pathology file size could range as high as 10-100x that of radiology. Furthermore, processing, ingesting, and storing WSI files using the same storage solution demands more performance than most healthcare organizations have in place today. Many organizations leverage spinning disk (HDD) storage solutions or Hybrid (SSD and HDD) solutions for Enterprise Imaging storage. These legacy technology types can quickly become overrun with performance problems due to the variability, scale, and volume of DICOM and non-DICOM data now expanding into Digital pathology and beyond (genomics, etc).

According to another article by ***J Pathol Inform,*** large pathology centers may produce over 1 million digital slides (~1 petabyte) of uncompressed data annually.[1]

Organizations must consider the impact of mixed file sizes from Radiology and Cardiology, the DICOM pathology conversion overhead, routing all these data sets across the enterprise, and ingest and storage capabilities to manage it all. Our ***previous white paper*** revealed that when the Dicom Systems Unifier® platform runs on Pure Storage® hardware, the combined solution can migrate more than 800,000* radiology imaging studies daily.

This white paper focuses on the routing functionality of the Dicom Systems Unifier® platform and the management of the digitized pathology DICOM slides on Pure Storage. We also measured the Ingest performance of WSIs writing to Pure Storage to measure data storage's impact. In the testing, we measured throughput, slide transmission count, and latency, broken into WSI routing and ingestion. The ***industry average size of a digital pathology slide***, 1.3 GB, was used for this whitepaper to calculate results. It is important to note that in some cases, like radical prostatectomies, an uncompressed 20x WSI can require nearly 30 GB of storage. At the recommended 40x magnification, the same file size can increase to 100 GB.

Routing Statistical Summary

Using the industry average slide size of 1.3 GB, the collaboration between Dicom Systems and Pure Storage resulted in an optimized solution that efficiently routed a maximum theoretical volume of 85,735 digital pathology slides daily.

- **85,735 digital pathology slides per day**
- **3,572 digital pathology slides per hour**
- **59.5 digital pathology slides per minute**

Routing Conclusion:

The Unifier® platform DICOM router's performance effectively scales with the system resources exhibiting a maximum routing throughput for 3DHistech, reaching up to 1.29 GB/s or 85,735 slides per day. At the same time, the performance of routing studies generated by Leica can reach up to .78 GB/s or around 51,840 slides routed daily. The solution leverages optimized DICOM routing, up to 20 Gbps aggregate network speed for incoming traffic, and up to 20 Gbps for outgoing, high-speed storage, and parallel processing capabilities. We add additional network capabilities because we can max out the network first before we utilize storage or other resources if we use only the industry standard maximum of 10 Gbps.

Ingest Statistical Summary

Using the industry average slide size of 1.3 GB, the collaboration between Dicom Systems and Pure Storage resulted in an optimized solution that efficiently ingests a maximum volume of 36,952 digital pathology slides daily.

- **36,952.6 digital pathology slides per day**
- **1,539.7 digital pathology slides per hour**
- **25.6 digital pathology slides per minute**

Ingest Conclusion:

The Unifier® Vendor Neutral Archive (VNA) ingest performance effectively scales with the system resources with maximum ingest throughput for 3DHistech, reaching up to 0.553 GB/s or 36,753 slides per day. At the same time, the performance of routing studies generated by Leica can reach up to 0.556 GB/s or around 36,952 slides routed daily. The solution leverages optimized DICOM ingest, high-speed storage, and parallel processing capabilities.

Our testing concluded that Dicom Systems, in combination with Pure Storage, ***can effortlessly route ~85,735 slides daily***, equivalent to managing the traffic of 29 3DHistech Pannoramic® 1000 scanners sending slides 24/7,* and ingest ~36,953 slides daily, equivalent to 13.5x the annual volume of a single, large pathology organization.**

* Using 3DHistech's Pannoramic® 1000's statistics of 30 seconds per slide, that equates to a speed of 2,880 slides per day if the scanner ran 24x7.

**According to J Pathol Inform, large pathology centers may produce over 1M digital slides (~1 petabyte) of uncompressed data annually, which equates to 2739 slides per day. Digital pathology volume and scanner deployments vary by organization. Actual workload will vary based on scanner speed and uptime, number of scanners, staffing, and infrastructure capabilities.

Methodology and Details Results

Using the Unifier® platform's DICOM routing and VNA functionality paired with Pure Storage as a single solution, we measured the impact of routing and ingesting DICOM whole-slide images. The research addressed varying image sizes and scanner manufacturers, workflow scenarios, and hardware resource allocation. The Unifier® platform's standard storage configuration was utilized for this test.

For testing, we have defined terms as the following:

- **Case** refers to a patient's sample or specimen that a pathologist analyzes. A case may consist of one or more slides.
- **Slide** is the physical glass slide that contains the tissue sample. WSI digitizes these glass slides into digital image files.
- **Tiles** are the smaller, individual image segments that comprise the whole slide image. Whole slide scanners capture the slide by imaging it tile-by-tile or in a line-scanning fashion and then stitching these tiles together into a single high-resolution digital image.
- **Routing** directs DICOM data, such as medical images (MRI study or Digital pathology case) and associated information, to appropriate destinations within healthcare networks. During this phase, network load and overhead of the DICOM conversion can become an issue for organizations.
- **Ingest** is the processing and storing of routed data for a VNA or chosen data platform. During this phase, the back-and-forth communication between the VNA application and the storage layer can become a bottleneck for organizations.

PURE STORAGE

Pure Storage enterprise imaging solutions uniquely support the initiatives driving modern healthcare and patient care. The testing in this white paper involved FlashArray//X™ and FlashArray//C™; however, the Pure portfolio extends further with:

FlashBlade® increases radiologists' and technicians' productivity with low latency while optimizing for AI workloads.

FlashArray//E™ and FlashBlade//E™ bring an all-flash data center to life, enabling a cloud operating model to drive simplicity, scalability, and better FinOps economics than legacy storage and cloud solutions.

Evergreen//One™ storage-as-a-service solution that provides the benefits of the public cloud with the security and performance of on-premises all-flash storage.

Summary of routing tests for FlashArray //X90R3 - SAN:

Size of objects	Speed GB/s	DICOM Files/s	Tiles/s
Video	0.978	0.78	n/a
Leica	0.777	4	87,513
3DHistech	1.289	78	19,031

WSI Routing Results:

The Unifier® DICOM router's performance scales with system resources, at maximum throughput for 3DHistech reaching up to 1.29 GB/s. At the same time, the performance of routing studies generated by Leica can reach up to .78 GB/s.

The critical factors for performance and scalability are storage space and configuration, network, and computing. The Pure Storage solution required no tuning when we ran the various configuration scenarios. It was an off-the-shelf deployment with standard VMware configuration without maximizing the network throughput limit in VMware configuration.

Summary of ingest tests for Flasharray //C60 - NFS:

Size of objects	Speed GB/s	DICOM Files/s	Tiles/s
Video	0.546	0.43	n/a
Leica	0.556	2.9	60,751
3DHistech	0.553	32	7,868

WSI Ingest Results:

The Unifier® DICOM VNA performance scales with system resources, reaching maximum throughput for Leica WSI samples of up to 0.556 GB/s. At the same time, the performance of routing studies generated by 3DHistech can reach up to .553 GB/s.

Like routing testing, the critical factors for performance and scalability are storage space and configuration, network, and computing. The Pure Storage solution required no tuning when we ran the various configuration scenarios. It was an off-the-shelf deployment with standard VMware configuration without maximizing the network throughput limit in VMware configuration.

TYPES OF STUDIES TESTING

Our testing data sets were fully anonymized pathology cases from 3 vendors:

Pathology Video File: iBox cameras and a video management system
Case Size: 10 GB
DICOM Files per case: 8
Format: MPEG-4 AVC/H.264 High Profile / Level 4.1
SOP Class: 1.2.840.10008.5.1.4.1.1.77.1.1.1, VideoEndoscopicImageStorage

DICOM Pathology WSI Sample 1:
Leica Biosystems GT450 DX
Case Size: 1139 MB
DICOM Files per case: 6
Total number of tile images: 125,223
Format: JPEG Baseline
SOP Class: 1.2.840.10008.5.1.4.1.1.77.1.6, VLWholeSlideMicroscopyImageStorage

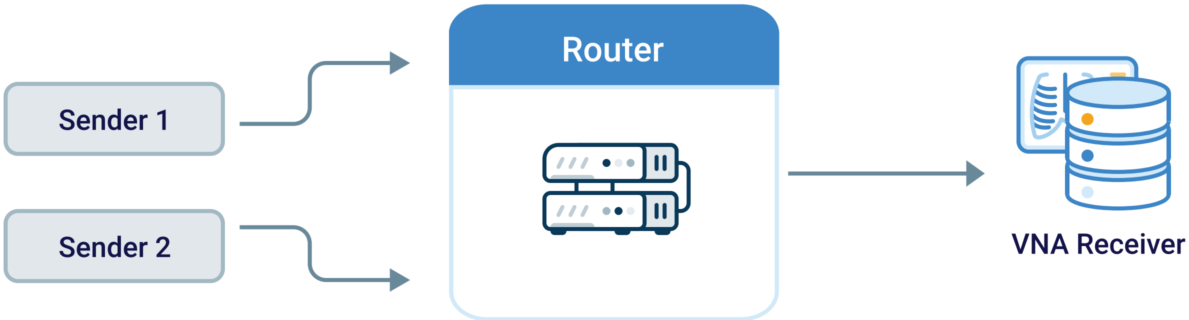
The Leica Aperio GT 450 can scan a 15 mm × 15 mm tissue area on a pathology slide at 40× magnification in just 32 seconds; this scanner boasts a throughput of 81 slides per hour. Such rapid processing enables efficient slide handling in pathology labs, accelerating turnaround times for slide analysis.

DICOM Pathology WSI Sample 2:
3DHistech Pannoramic 1000
Case Size: 218 MB
DICOM Files per case: 13
Total number of tiles: 3143
Format: JPEG Baseline
SOP Class: 1.2.840.10008.5.1.4.1.1.77.1.6, VLWholeSlideMicroscopyImageStorage

The 3DHISTECH Pannoramic 1000 scanner is the fastest whole-slide scanner on the market. It can scan up to 100 slides per hour, achieving this speed with specific settings like a 20x objective/40x optical equivalent magnification, 0.24 µm/pixel resolution, 7500 µm focus point distance, and an average sample size of 15×15 mm.

PART 1 - ROUTING

FlashArray//X90-R3 & Unifier VNA (set to router configuration)



Test 1.1 - Routing Pathology Video: iBOX

This test, which ran for at least 10 hours, measured the performance of a pathology endoscopic video made of MPEG-4 Video files.

- The Unifier VNA was configured as a mock receiver (as a black hole), minimizing delay to ensure the VNA ingest speed did not affect or limit the testing. PostgreSQL DB, storage, and index files were on separate volumes served off a single FlashArray//X90-R3. Storage had FIFO mode enabled, a 1TB quota set, and the standard type of Unifier® storage. Each thread had a separate block of storage and index volume configured on the backend.
- The receiver has a 2 TB storage allocation on the FlashArray//X90-R3. PostgreSQL DB, cache storage, and index files are on separate volumes; each thread has a separate storage block, and index disks are configured on the backend.
- Senders 1 and 2 are used for “feeding” the router utilizing both 10Gbps network interfaces installed. The router sends studies to VNA using both 10Gbps network cards. The router has 4 network cards installed, 2 for input and 2 for output.
- Both senders use a similar pack of specially prepared studies. The maximum number of threads for sending is 6, and the maximum for routing is 50.

SUMMARY OF RESULTS

.978 GB/s

The maximum routing speed achieved was .978 GB/s for MPEG-4 video routing.

Detailed Summary

- Storage utilization is ~87% (max queue size of 64)
- CPU utilization ~ is 11%
- Maximum network utilization is ~36% of the total throughput

Our testing confirmed that routing speed is limited mainly by the network card and storage speed. This is possible due to high cache utilization and the availability of sufficient RAM for caching.

The minimum hardware requirements for this type of workflow are 8 CPU cores, 1 6GB RAM (32 GB recommended), a 10 GB/s network card, and fast storage. 2 TB disk storage can store about 30 minutes of cache.

+ **Further details on Hardware Specifications are in Appendix A. Details on Test 1.1 are in Appendix B.**

Test 1.2 - Routing DICOM Pathology WSI Sample 1: Leica

This test, which ran for at least 10 hours, measured the performance of WSI DICOM image routing with a pathology case from a Leica scanner.

- The VNA Receiver was configured as a mock receiver (as a black hole) to measure routing performance alone. The standard type of Unifier® storage configuration was selected, FIFO mode enabled, and a 1 TB quota set.
- The receiver has a 2 TB storage allocation on the FlashArray//X90-R3. PostgreSQL DB, cache storage, and index files are on separate volumes; each thread has a separate storage block, and index disks are configured on the backend.
- Senders 1 and 2 are used for "feeding" the Router utilizing both 10 Gbps network interfaces installed. The router sends studies to VNA using both 10 Gbps network cards. The router has 4 network cards installed, 2 for input and 2 for output.
- Both senders use a similar pack of specially prepared studies. The maximum number of threads for sending is 6, and the maximum for routing is 50.

SUMMARY OF RESULTS

0.777 GB/s
1 slide every 1.63 seconds

The maximum routing speed was 0.777 GB/s, or approximately 1 slide every 2 seconds, assuming 1.3 GB per slide on average.

Detailed Summary

- Storage Utilization is 79% (maximum queue size ~42)
- CPU utilization ~16%
- Maximum network utilization ~30% of total throughput

The testing showed that disk writing speed is a limiting factor, which is possible due to high cache utilization and the availability of sufficient RAM for caching.

The minimum hardware requirements for this type of workflow are 16 CPU cores, 64GB RAM (96 GB recommended), a 10 GB/s network card, and fast storage. Two TB disk storage can store about 40 minutes of cache.

+ Further details on Hardware Specifications are in Appendix A. Details on Test 1.2 are in Appendix C.

Test 1.3 - Routing DICOM Pathology WSI Sample 2: 3DHistech

The test, which ran for at least 10 hours, measured the performance of WSI DICOM image routing with a pathology case from a 3DHistech scanner.

- The VNA Receiver was configured as a mock receiver (as a black hole) to measure routing performance alone. The standard type of Unifier® storage configuration was selected, FIFO mode enabled, and a 1 TB quota set.
- The receiver has a total of 2 TB storage allocation on the FlashArray//X90-R3. PostgreSQL DB, cache storage, and index files are on separate volumes; each thread has a separate storage block, and index disks are configured on the backend.
- Senders 1 and 2 are used for “feeding” the Router utilizing both 10 Gbps network interfaces installed. The router sends studies to VNA using both 10 Gbps network cards. The router has 4 network cards installed, 2 for input and 2 for output.
- Both senders use a similar pack of specially prepared studies. The maximum number of threads for sending is 6, and the maximum for routing is 50.

SUMMARY OF RESULTS

1.289 GB/s
1.008 every second

The maximum routing speed was 1.289 GB/s, approximately 1 slide per second, assuming a 1.3 GB per average slide.

Detailed Summary

- Storage disk utilization is 79% maximum (queue size ~42)
- CPU utilization is about 22%
- Network utilization ~30% of total throughput

Performance bottlenecks again were RAM and disk speeds.

The minimum hardware requirements for this type of workflow are 16 CPU cores, 48 GB RAM (96 GB recommended), 10 Gb/s network card, and fast storage. 2 TB storage will keep the cache for ~24 minutes.

+ **Further details on Hardware Specifications are in Appendix A. Details on Test 1.3 are in Appendix D.**

PART 2 - INGEST

FlashArray//C60-R2 & Unifier VNA

Test 2.1 - Ingest Pathology Video: iBOX

The test aims to measure the ingest performance of a pathology endoscopic video made of an MPEG-4 Video file. The test duration was 21 hours (39+TB of data processed and stored).

- The VNA Receiver was configured as a regular receiver with the NFS disk for file storage. The type of Unifier® storage is set to VNA. A quota was not set with the file data stored on FA//C60 with FIFO mode disabled.
- PostgreSQL DB, cache storage, and index files are on separate volumes; each thread has a separate storage block, and index disks are configured on the backend.
- Senders 1, 2, and 3 are used for "feeding" the primary VNA Receiver. The router has both 10 Gbps network interfaces installed and sends studies to VNA using both 10 Gbps network cards. It has 4 network cards installed, 3 for input and 1 for output.
- The three senders use a similar pack of specially prepared studies. The maximum number of threads for sending is 3.

SUMMARY OF RESULTS

0.533GB/s

Maximum ingest speed – 0.533 GB/s.

During the test phase, 4,048 videos, 32,384 files, and 39.64 TB were sent to the VNA.

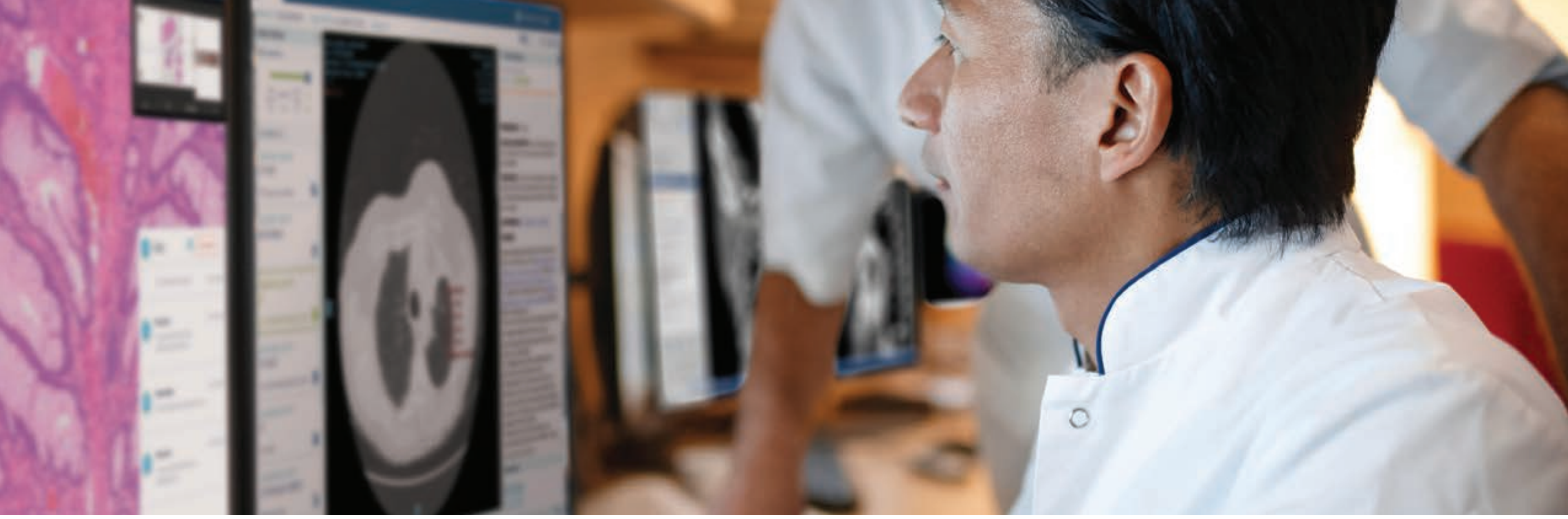
Detailed Summary

- Storage disk utilization cannot be measured from an OS perspective as it is a network-based NFS mount point.
- CPU utilization is about 10%.
- Network utilization is ~45% maximum on a single network card.

The routing speed is limited mainly by storage writing speed over the network and NFS mount.

The minimum hardware requirements for this type of workflow are 8 CPU cores, 32 GB RAM, and a 10 Gb/s network card. The performance bottleneck is network speed.

+ **Further details on Hardware Specifications are in Appendix A. Details on Test 2.1 are in Appendix E.**



Test 2.2 - Ingest DICOM Pathology WSI Sample 2: Leica

The test, which ran for at least 10 hours, measured the performance of WSI DICOM images ingested with a pathology case from a Leica scanner..

- The VNA Receiver was configured as a regular receiver with the NFS disk for file storage. The Unifier® storage type is set to VNA. A quota is not set with the file data stored on FA//C60 with FIFO mode disabled.
- PostgreSQL DB, cache storage, and index files are on separate volumes; each thread has a separate storage block, and index disks are configured on the backend.
- Senders 1, 2, and 3 are used for "feeding" the primary VNA Receiver. The router has both 10Gbps network interfaces installed and sends studies to VNA using both 10 Gbps network cards. It has 4 network cards installed, 3 for input and 1 for output.
- The three senders use a similar pack of specially prepared studies. The maximum number of threads for sending is 3.

SUMMARY OF RESULTS

0.539 GB/s

Max ingest speed – 0.539 GB/s. Resulting as 0.41 slides/s, 2.9 files/s, 60751 tile/s.

During the test phase, the following data was sent to the VNA: 13,653 slides, 81,920 files, and 14.84 TB of data was ingested.

Detailed Summary

- Storage disk utilization cannot be measured from an OS perspective as it is a network-based NFS mount point.
- CPU utilization was approximately 10%.
- Network utilization is ~40% maximum on a single network card.

The routing speed is limited mainly by storage writing speed over the network and NFS mount.

The minimum hardware requirements for this type of workflow are 8 CPU cores, 32 GB RAM, and a 10 Gb/s network card. The performance bottleneck is network speed.

+ Further details on Hardware Specifications are in Appendix A. Details on Test 2.2 are in Appendix F.

Test 2.3 - Ingest DICOM Pathology WSI Sample 2: 3DHistech

The test lasted 12 hours (22+TB). It aimed to measure the performance of WSI Storage of DICOM studies with pathology studies from a Leica scanner.

- The VNA Receiver was configured as a regular receiver with the NFS volume for file storage. The type of Unifier® storage is set to VNA, and a quota is not set with the file data stored on FA//C60 with FIFO mode disabled.
- PostgreSQL DB, cache storage, and index files are on separate volumes; each thread has a separate storage block, and index disks are configured on the backend.
- Senders 1, 2, and 3 are used for "feeding" the primary VNA Receiver. The router has both 10Gbps network interfaces installed and sends studies to VNA using both 10Gbps network cards. It has 4 network cards installed, 3 for input and 1 for output.
- The three senders use a similar pack of specially prepared studies. The maximum number of threads for sending is 3.

SUMMARY OF RESULTS

0.521 GB/s

Max ingest speed – 0.521 GB/s. Resulting as 0.4 slides/s, 32 images/s, 7,868 tiles/s.

During the test phase, the following data were sent to the VNA: 109603 slides, 1,424,839 files, and 22.28 TB of data.

Detailed Summary

- Storage disk utilization cannot be measured from an OS perspective as it is a network-based NFS mount point.
- CPU utilization is about 14 %
- Network utilization is ~45% maximum on a single network card.

The ingest speed is limited mainly by storage writing speed over the network and NFS mount.

The minimum hardware requirements for this type of workflow are 8 CPU cores, 32 GB RAM, and a 10 Gb/s network card. The performance bottleneck is network speed.

+ **Further details on Hardware Specifications are in Appendix A. Details on Test 2.3 are in Appendix G.**

For more information call
415.684.8790 or visit **dcmsys.com**

Reference
1 Ashman K, Huimin Zhuge, Shanley E, et al. Whole slide image data utilization informed by digital diagnosis patterns. Journal of pathology informatics. 2022;13:100113-100113.

PURESTORAGE® dicom systems