komprise | NewYork-Presbyterian

KDX
THE KOMPRISE DATA EXPERIENCE

CASE STUDY

NewYork-Presbyterian Achieves 96% Savings and 10x Faster AI Data Ingestion with Komprise

Overview

NewYork-Presbyterian is a leading academic healthcare system consistently ranked among the top hospitals nationally by U.S. News & World Report. The organization consists of 10 hospital campuses and more than 200 clinics in the New York metropolitan region. NewYork-Presbyterian is undergoing several modernization initiatives spanning digital health and AI, with digital pathology a pioneering use case. The IT infrastructure team is heavily involved in these initiatives, from modernizing storage to delivering the best technologies for AI data preparation and ingestion.

IT Faces Sudden Focus on AI Data Management

Zhani Pellumbi, the senior technical architect of IT infrastructure at NewYork-Presbyterian, is responsible for creating and maintaining a high-performing, resilient hybrid cloud environment for hosting clinical AI applications and managing petabytes of unstructured clinical data.

This data, captured at point-of-care such as from cameras on ambulances and in the ER, requires long-term storage to comply with industry regulations. "The number of images that we're ingesting which are part of patients' medical history is exploding. Most of it is already imaged and you can't compress or dedupe it."

Key Stats

Industry: Healthcare

Storage: Pure Storage, AWS S3

Data Analyzed and Ingested: ~3PB/yr, 28GB of new data ingested every 5 minutes

Use Cases: AI Data Ingestion, AI Data Workflows

Results: 96% lower cloud storage costs, 10x faster data ingestion, reduced security exposure in cloud, easier research access to clinical files.

Beyond the fundamental storage and backup challenge, a high priority project in 2025 entailed supporting a digital pathology program to improve patient diagnosis using cloud-based AI. The pathology clinical team scans tissue slides around the clock, generating massive image files, which amount to more than 2PB of new data every year.

"We have large datasets for unstructured imaging, and needed a way to ingest it, store it, query it and move it, while being cost effective and meeting the strict performance requirements of our stakeholders," Pellumbi said.

Top Challenges:

- Medical images need to be retained on-premises for decades.

- Multi-petabytes of data; each pathology slide can be 50+ GB.

- The cloud data transfer tool copies all data on the source, driving up storage costs.

- Initiative to leverage cloud-based AI, while protecting PHI.

How it Works:

- Pathologists access results in PathAI and verify and/or adjust findings as needed.

- After 30 days, the copies are deleted according to policy.

- The PathAI application can call a Komprise API to reingest older slides into AWS S3, as needed by pathology.

- Meanwhile, the original data stays safely on-premises for long-term retention and compliance.

Komprise: Automated, Cost-Efficient AI Data Workflows

The health system discovered that the Komprise Intelligent Data Management platform for unstructured data management could efficiently execute the steps of this detailed AI ingestion process and cut cloud storage costs. The technology behind this, Komprise Intelligent AI Ingest, automates the discovery, classification, tagging, and ingestion of unstructured data across storage silos to AI pipelines.

 "This is the first of its kind project at the health system," Pellumbi said. Instead of lab technicians looking at slides under the microscope, counting mitosis or measuring tumors, the PathAI software does the core analysis, which is verified by a human who can easily view the slides on large color monitors. "This means less mistakes and a faster turnaround," he added.

"

"Most of our data is unstructured, untagged and nobody knows what's in there or how frequently it's being accessed. The storage vendors haven't provided good tooling to look inside that data set. Komprise gives us visibility into this data, it can tag that data, and then have it ready for AI. Being able to move just the right amount of data where you need it at those speeds is where Komprise really shines for us."

— *Zhani Pellumbi*,
Senior technical architect of IT infrastructure, NewYork-Presbyterian

Benefits of Komprise

With Komprise, AWS S3 and PathAI, NewYork-Presbyterian is finding success with AI at the point of care in a bold new initiative to improve treatment and outcomes. The solution has allowed IT to operationalize AI with great economics, speed and accuracy for their constituents.

Cost Savings: IT reduced cloud costs by 96% using the Komprise automated workflow that curates a small subset of files and then deletes cloud copies after 30 days. This approach pared down AWS storage from 1PB to a rolling 33TB.

Precise AI Ingestion: Unlike copy-and-sync tools that copy everything, Komprise curates and only sends the newest data to the cloud , at speeds 10x faster than previous data transfer methods in use at NewYork-Presbyterian.

Accelerated AI momentum: Komprise is paying for itself through the cost savings achieved with the surgical cloud copy and delete workflow. "With this project we could fund it quickly because the ROI is tied into our spend and it's not being funded as an infrastructure tool," Pellumbi explains.

Granular, metadata-based search and tagging: Researchers can find relevant data by querying the Komprise Global Metadatabase without having to run AI again. This is possible through policy-based data tagging of curated files. Komprise metadata tags persist as data moves between storage platforms so that users can always locate files and queries.

Storage-agnostic scalability: Until Komprise, there hasn't been a viable way to filter and move the organization's unstructured data. The Komprise elastic scale architecture handles petabytes of diverse data and moves data transparently, without stubs. "We don't need the original NAS to access our data if we want to recall it," Pellumbi said. "File integrity across storage platforms is important for our present and future needs."

Lower security risks: "Our information security director was tremendously delighted when we reduced our cloud storage footprint from 1 petabyte to 30 terabytes, because it's less exposure," Pellumbi says.

Fueling clinical AI success: By using Komprise to ingest the right data to PathAI, NY Presbyterian's health system can integrate machine learning algorithms with digital pathology systems. This improves precision in detecting and characterizing disease features, such as complex biomarkers, while also delivering assessments to support personalized care, especially in oncology and immuno-oncology.

Setting the Stage for Clinical AI Practice

Pellumbi's team is planning to use Komprise to enrich metadata for self-service search, integrate with Snowflake to merge unstructured data queries with structured data, and expand the AI ingestion workflow to new specialties such as radiology. NewYork-Presbyterian's foray into AI data preparation and ingestion of pathology data is a groundbreaking step into a new era of medicine.