

# HM Health Solutions protects next-generation healthcare services with an always-on data lake

## Company overview

HM Health Solutions delivers business solutions that empower health plan payers to run their organizations efficiently in a competitive and ever-changing market. Offering cutting-edge technology and deep industry knowledge, HM Health Solutions is partnered with 13 health plans serving 10 million members. Headquartered in Pittsburgh, Pennsylvania, HM Health Solutions employs more than 3,500 people.

## Challenge

As it set out to build its first data lake, HM Health Solutions aimed to protect the platform from disaster 24/7. How could the company ensure that its cutting-edge Hadoop cluster offered 24/7 availability?

## Solution

HM Health Solutions deployed Cirata to continuously replicate large, fast-changing HDFS and Apache Hive data sets between its primary data lake and a disaster recovery environment.

## Results

- Enables seamless business continuity in the event of an outage, reducing operational disruption
- Ensures near-zero data loss in a recovery scenario, keeping valuable healthcare data protected
- Offers a scalable solution able to handle predicted 30% year-on-year increase in data volumes

To empower its affiliates to build data-driven healthcare services, HM Health Solutions decided to create an enterprise data lake—and protecting the new platform from disaster was a top priority. By deploying Cirata to continuously replicate data to a recovery environment, HM Health Solutions ensures near-instant failover and zero data loss in the event of an outage at its primary data center.

## Preparing for a data-driven future

In the United States, health plans and providers face tough challenges. The combination of an aging population and a constantly changing regulatory environment means that cost-effective delivery is more important than ever. And even though margins are tightening, delivering high-quality care experiences is essential to drive membership growth and encourage long-term member retention.

HM Health Solutions recognized that a new generation of machine learning (ML) and artificial intelligence (AI) capabilities could empower its healthcare affiliates to harness their data to solve these challenges.

Steven Swartzlander, Lead Architect at HM Health Solutions, explains: “The ability to perform analytics at scale opens up a world of opportunities—from modeling the risk of hospital infections to validating that procedures have been billed correctly. AI and ML capabilities make it possible to enhance decision making processes with insights that would have been far too complex and costly to obtain through manual analysis.”

Rajesh Nambadi, IT Manager at HM Health Solutions, continues: “In recent years, we’ve seen an increasing volume and variety of analytics requests from our health plan and provider affiliates. We knew that our existing big data analytics environment, based on a Teradata appliance, would soon be unable to accommodate the growing appetite for data-driven services.”

To support informatics and data science teams across its affiliate network, HM Health Solutions decided to build a cutting-edge data lake, based on a Hortonworks Data Platform Hadoop cluster. The aim was to create a one-stop shop for analytics workloads, with the scalability to accommodate growing data volumes in the long-term.

“In the past, we did not have a dedicated disaster recovery [DR] capability for our analytics platform, as it captured data from line-of-business databases that were already protected by DR strategies elsewhere in the business,” Steven Swartzlander continues. “Because the new Hadoop cluster was intended to become a critical data platform, we looked for a way to deliver round-the-clock availability with full protection for Hadoop Distributed File System [HDFS] and Apache Hive data.”

## Protecting key analytics services

After meeting with industry analysts and its Hadoop vendor, HM Health Solutions identified two options for replicating its data to a secondary DR site.

“One of our options was to use Apache Falcon to move data in batches via the DistCp tool,” recalls Steven Swartzlander. “However, we quickly realized that this approach would not allow us to keep data in our DR and production environments synchronized, which increased the risk of data loss in a recovery scenario. This approach would also have required us to knit together a giant patchwork of custom scripts—demanding significant ongoing investment in maintenance, and potentially even wholesale re-development if the tool were removed from the Hadoop distribution.”

To achieve its DR and high-availability goals, HM Health Solutions selected Cirata to continuously replicate HDFS and Hive data between both sites.

“One of the things that impressed us most about Cirata was its ability to resolve the data consistency challenges of replicating large amounts of data between active Hadoop clusters,” explains Steven Swartzlander. “Because Cirata enforces consistency between each environment automatically, there’s no need for manual intervention from our team—enabling us to focus on value-added development work.”

By configuring Cirata for bidirectional replication, HM Health Solutions’ DR environment could be used for production workloads, effectively doubling its compute capacity without compromising on its DR capabilities.

Working together with an expert team from Cirata, HM Health Solutions deployed a test environment in its secondary data center.

"Collaborating with the Cirata team during the testing process was extremely valuable, as our experience to DR prior to that exercise had predominantly been storage replication alone," comments Steven Swartzlander. "Throughout the testing and implementation process, Cirata provided us with all the support and guidance we needed to ensure our deployment was a success. Whenever we had a technical question, we knew that the answer we needed was only a phone call away."

Following a successful proof-of-concept exercise, HM Health Solutions configured Cirata to replicate its production environment to an active cluster at its DR site.

"Overall, our implementation process went extremely smoothly, and we only had to make minor changes to the standard configuration of Cirata," says Steven Swartzlander. "In particular, our Linux administrators wanted to ensure that replicating large volumes of data between our two data centers wouldn't overwhelm the default network link, which uses a lower bandwidth channel. By making some straightforward changes in Cirata, we successfully redirected the traffic to a separate network interface with higher bandwidth."

## Enabling healthcare innovation

With Cirata ensuring consistent data at its production and DR cluster, HM Health Solutions is achieving its goal of providing advanced AI and ML capabilities to stakeholders across the organization, 24/7.

"If the Hadoop platform were to go offline, the dedicated analytics teams who rely on it would be unable to work, resulting in a significant loss of productivity," explains Steven Swartzlander. "By using Cirata to replicate our data continuously to the DR site, we have practically no risk of data loss if we suffer an outage at the primary cluster. And because our DR cluster is active 24/7, we can fail over almost instantly without our users even realizing. When you consider that our recovery time objective [RTO] is six hours, our ability to deliver near-zero RTO with Cirata is an outstanding achievement."

Today, the organization holds around 300 TB of data on each of its clusters—and looking ahead, HM Health Solutions predicts 30 percent year-on-year data growth. Because the Cirata platform is designed for large, fast changing data sets, HM Health Solutions can continue to scale out its DR environment with confidence.

"One of our bigger users of the cluster at the moment is combining optical character recognition and natural language processing to ingest and process insurance claim data," adds Steven Swartzlander. "This AI solution can identify cases where treatments weren't coded correctly, enabling the organization to apply for reimbursement for the government, recover lost revenue, and safeguard its margins. Solutions like these play a significant role in data growth on the platform—and with Cirata, we know we can keep scaling out without any impact on our DR regime."

Looking to the future, HM Health Solutions is exploring opportunities to harness Cirata to further enhance its approach to analytics. By enabling bidirectional replication between the DR site and the production cluster, changes at that site would be reflected in the production site. In this configuration, the secondary environment could be used for production workloads, effectively doubling HM Health Solutions' compute capacity without compromising on its DR capabilities.

"With Cirata, we've gained the ability to protect our Hadoop clusters from disaster without the cost and complexity of batch-based approaches such as DistCp," comments Steven Swartzlander. "We know that Cirata is equally capable of replicating data to the cloud. Going forward, we'd be interested in using this ability to enable cloudburst-style capabilities to accommodate peaks in analytics demand." Steven Swartzlander concludes: "Today, we can offer our data science and informatics teams round-the-clock access to analytics at scale. We see that data-driven services are going to be crucial to deliver high-quality patient experiences cost-effectively—and by partnering with Cirata, we can protect this vital data 24/7."

