

Sanlam ensures high availability with a disaster recovery implementation that supports near-zero recovery point objectives

Objective

Sanlam's IT operations group needed to ensure high availability for their Cloudera Data Platform (CDP) implementation. To achieve this objective, they determined they needed to introduce a Disaster Recovery (DR) environment that could quickly take over if a failure or outage occurred in the primary production environment.

Challenge

Implementing a DR environment that can quickly take over in the event of a failure requires the data to be kept consistent with the data in the production environment. How could Sanlam replicate the actively-changing production data without impacting performance in order to meet the high availability objectives for their CDP implementation?

Solution

Sanlam selected Cirata Data Migrator both to perform the initial migration and the ongoing replication of changes from production to the DR environment. Data Migrator replicates changes as they occur in production so that data is kept consistent, enabling Sanlam to support a near-zero recovery time objective (RTO) and recovery point objective (RPO).

Business impact

- Original data migration of 70TB of data performed with no business disruption
- Ongoing replication of changes to production data as they occur (20TB per week)
- No disruption or impact to existing production environment
- Established DR environment that supports near-zero RTO and RPO
- DR environment provides additional capacity for more analytics, AI, and ML processing
- Enables Sanlam to easily move data to other sources as well, such as for future cloud requirements

Ongoing replication of data changes as they occur enables near-zero RPO and RTO

Customer profile

Sanlam was established as a life insurance company in South Africa but has since transformed into a diversified financial services group operating across Africa, India, and selected emerging and developed markets, with listings on the Johannesburg, A2X, and Namibian stock exchanges.

Sanlam has been creating value for stakeholders since 1918 — for more than 100 years. Digital capabilities are critical to transforming their business for client-centric growth. The Group focuses on innovation across their products and services, distribution channels, and back-office processes. Sanlam is on a journey to digital business transformation by digitally optimizing their traditional business to a more efficient tech enabled provider, and investing in new capabilities and assets to build a disruptive 3.0 platform (see Figure 1).

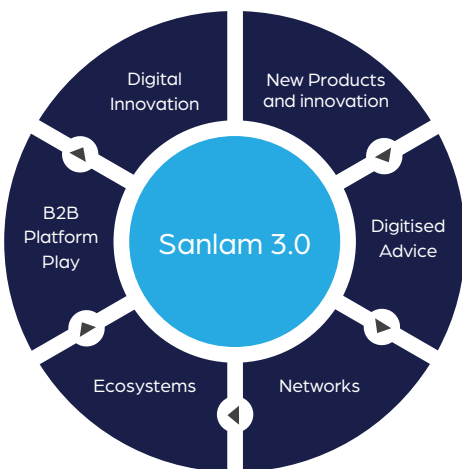


Figure 1: Sanlam 3.0 platform

Objectives

One of the key pillars in Sanlam's approach to digital innovation includes implementing a Group-wide business intelligence platform through big data and enhanced data analytics.

Sanlam recognizes disruptive threats, cyber risks, and various other risks associated with digitized information. To mitigate the risks, Sanlam has established an objective to ensure high availability of their critical business intelligence platforms, which includes their CDP implementation. Specific objectives include:

- Implement a DR environment that can quickly take over if a failure or outage occurs to the primary CDP environment
- Replicate data changes to keep DR environment consistent with primary environment
- Support objectives for near-zero RPO and RTO

Challenges

Meeting the high availability objectives that Sanlam established for their CDP implementation meant that data in the DR environment needed to be kept consistent with that in production. Data is continuously being changed in production — either new data ingested, or existing data updated — so Sanlam required a solution that could replicate the data changes as they occurred production to the DR environment.

Sanlam investigated several tools, such as Cloudera Replication Manager and DistCp (Distributed Copy) — an open-source tool provided with their Hadoop distribution. However, these tools do not support replication of changes as they occur. Instead, users must create schedules to periodically replicate data incrementally. In the event of a failure, any changes made since the prior replication run will be lost, impacting both the RPO and RTO. Furthermore, these DistCp-based solutions run as standard MapReduce jobs competing for resources with other processes, which can impact production system performance.

Sanlam needed a solution that could replicate the actively-changing production data in near-real time without impacting performance so they could meet their high availability objectives.

Solution

Following a thorough technical evaluation, Sanlam selected Cirata Data Migrator to automate the migration and ongoing replication of data between their primary and secondary (DR) environments.

Data Migrator is a safe and reliable cloud migration solution that automates the migration and replication of Hadoop data and Hive metadata to the cloud or between data centers. Data Migrator deployment is performed in minutes and requires no changes to applications or business operations. Migrations and replication of any scale can begin immediately and be performed while the source data is under active change, without requiring any production system downtime or business disruption. Data Migrator is the ideal solution for Sanlam and enables them to support near-zero recovery time objective (RTO) and recovery point objective (RPO), which were critical objectives for their business intelligence platforms.

Jacques Joubert, Big Data Manager and Architect at Sanlam, stated: "We were impressed with Cirata's unique capabilities for real-time replication without disrupting our production environments. It was critical to us that data in our disaster recovery cluster is kept in sync with our primary cluster as closely as possible to enable a near-zero recovery point objective and recovery time objective. Cirata and its partners provided Sanlam with great support during the product evaluation and initial project deployment. They addressed our questions and concerns in an efficient and professional manner and delivered results within the timeline they promised."

Results

Sanlam worked closely with Cirata customer success to define the scope of the project and determine what would be required for a successful engagement. Following joint testing with the product, Data Migrator was deployed to an edge node of the source CDP cluster. Deployment was performed in minutes with no impact to current production operations, and Sanlam could begin using the product immediately.

Migrations are designed to be easy to configure and perform. During deployment, Data Migrator automatically identified the source HDFS cluster, so Sanlam only needed to define the target CDP environment and the datasets they wanted to migrate. Data Migrator provides full control over what data to migrate and what data to exclude.

Migrations can be automatically started and are performed with a single pass through the source storage system, eliminating the overhead of repeated scans while also supporting continuous replication of any ongoing changes from source to target, with zero disruption to the production environment.

The continuous replication of ongoing changes enables the near-zero RPO and RTO that was an important Sanlam objective. In addition, the DR environment can be used to provide additional capacity for more analytics, AI, and ML processing when needed.

While this initial project was to establish a high availability solution between CDP environments across two different data centers, the same Data Migrator implementation can be used to move data to other sources as well, including public cloud environments, which Sanlam expects to leverage in the future.

Results summary

- Original data migration of 70TB of data performed with no business disruption
- Ongoing replication of changes to production data as they occur (20TB per week)
- No disruption or impact to existing production environment
- Established DR environment that supports near-zero RTO and RPO
- DR environment provides additional capacity for more analytics, AI, and ML processing
- Data Migrator enables Sanlam to easily move data to other sources as well, such as for their future cloud requirements