Case Study

Stackable
your data, your platform

A Modern Data Platform for Artificial Intelligence

Discovery's Journey: Transforming Data Management in Financial Services

Discovery

Initial situation and challenges

Discovery, one of the largest South Africa-founded financial services organisations, provides most of its local business units with more than 60 data scientists with a **comprehensive data platform for advanced analytics and machine learning**.

Initially, their solution was a complex on-premises setup based on Cloudera, including HBase, YARN, HDFS, Kafka, NiFi and CDSW. However, the team was facing increasingly significant challenges with package management, version conflicts and failed attempts to run YARN with Kubernetes. HDFS was not optimised for their InfiniBand network, resulting in network overhead and poor hardware utilisation.

There was a need for a more flexible solution to make updates to frameworks less cumbersome and time-consuming. More agility and scalability was required, particularly for their data science and machine learning workloads. Following a thorough evaluation period, Discovery decided to transition to the Stackable Data Platform.

> *The Stackable Data Platform greatly simplifies the management of complex environments with numerous components thanks to its K8s operators. This significantly increases flexibility while reducing operational effort.*
>
> Nick Alexander, Systems Architect Discovery Health

Discovery

> *Working with Discovery's data engineering team was exceptional, and we were impressed by their technical expertise. By adopting our Stackable Data Platform, they created a comprehensive data and AI environment optimised down to the last detail. Discovery's forward-thinking approach is truly commendable.*
>
> Dr. Stefan Igel, COO at Stackable

Stackable

Solution

Discovery addressed its challenges by migrating to the Stackable Data Platform, which is built on **a modular, Kubernetes-native architecture**.

While the existing hardware could be reused by reallocating resources, the platform's software components have been **completely renewed or updated**: The team migrated from HDFS to VAST DataStore and configured **Apache HBase®** to run on top of it. Other systems such as Apache Hive™ and Impala have been replaced with **Trino**, which improved flexibility. **Apache Iceberg™** was introduced as the new storage format, together with enhanced privacy features. The team also transitioned from **Apache Spark™** 2 to Spark 3 to leverage improved performance, GPU acceleration, and the PySpark experience.

Finally, the data migration, including 400 TB of storage, was **completed over a weekend** with a prior dry-run to ensure system stability and minimal downtime.

Result and Successes

Following its migration to the Stackable Data Platform, Discovery has achieved a **significant increase in agility, scalability, and operational efficiency** across its data architecture.

Most business units, including those in insurance and healthcare, now run on **shared infrastructure that supports ETL, streaming, and machine learning workloads**. The shift to modular, operator-managed components has **reduced complexity** and staffing needs while **enabling the use of new tools** such as the AI compute engine Ray alongside Spark.

Discovery has **significantly improved the performance and flexibility of queries**, as well as transitioning to a future-proof data architecture. Enhanced data virtualisation provides seamless access to sources such as Oracle and Netezza. **Day-2 operations are made easy** with Stackable. Its open approach to system maintenance, monitoring, and observability allows for **modular component updates** and **integration with other third-party solutions**.

The result is a flexible, modern data stack that supports **performance and innovation**.

Highlights at a glance

High-performance AI data plaform to support over 50 data scientists in several business units

Minimized dependency on single vendor, enhancing flexibility and control over technology choices

Capability to incrementally upgrade to the latest versions of data products without disruption

Efficient data migration with minimal downtime

Greater adaptability and innovation through open-source solutions and customization

Trino as central part of the new solution allowing data virtualization across the organization

About Discovery

Discovery is a proudly South African-founded financial services organisation that operates in the healthcare, life insurance, short-term insurance, long-term savings, banking and wellness markets.

Since inception in 1992, Discovery has been guided by a clear core purpose - to make people healthier and to enhance and protect their lives. They have been able to do this by pioneering the shared-value insurance model, which delivers better health and value for clients, superior actuarial dynamics for the insurer, and a healthier society. The success of the model in the markets where Discovery operates has been testament to its importance to society.

More at www.discovery.co.za

About Stackable

With its open source data platform, Stackable stands for professional data sovereignty in the corporate context. The company attaches particular importance to data security, openness and transparency and offers first-class support at fair conditions.

The company, based in Wedel, Germany, was founded in 2020 out of the open source community. Stackable relies on open source code to support companies in dealing with big data. As an innovative and internationally active provider, Stackable puts the community at the center, promotes cooperation instead of competition and supports companies in the development of their data architecture.

More at www.stackable.tech