MANAGING DATA IN THE PRIVATE CLOUD ENABLES REMOTE ACCESS TO HEALTH DATA AND ANALYTICS FOR RESEARCH

IMPACT
- Faster and more reliable data available for remote workers
- Supporting external researchers and growing exponentially for remote users
- Centralized analysis for data located in multiple provinces

Dalhousie University is Atlantic Canada's leading research-intensive university and a driver of the region's intellectual, social and economic development. Located in the heart of Halifax, Nova Scotia, with an Agricultural Campus in Truro/Bible Hill, Dalhousie is a truly national and international university, with more than half of its 19,000 students coming from outside of the province. Dal's 6,000 faculty and staff foster a vibrant, purpose-driven community, that celebrated 200 years of academic excellence in 2018.

Within Dalhousie is Health Data Nova Scotia (HDNS), which provides access to de-identified, linkable health data in a secure, controlled environment. The HDNS team works in partnership with the Maritime SPOR SUPPORT Unit (MSSU) to bring together health policy decision makers, health care professionals, researchers, patients and caregivers to address priority health topics, create and strengthen data-sharing partnerships, and maximize the utility of data resources to support patient-oriented research.

HDNS strives to conduct research that is relevant and applicable to improve the health of those in the Maritimes and beyond. Key areas of focus include chronic disease prevention and management, primary health care, and mental health as well as healthcare planning, policies, medical research, treatments and outcomes. HDNS aims to provide researchers with better insights into specific diseases and treatment options and help people live healthier lives.

Challenge
HDNS was challenged with managing data and having information readily available, easily searchable and provided quickly to users, both local and remote. Primarily accessible to epidemiologists, the data also had to be available to all parties interested in using it for research purposes from across the province and the country. Managing data in a secure way for remote users was critical and building a platform to scale and support future users and use cases meant embarking on a big data journey.

HDNS supports a varied set of research projects, so delivering both data and analytics on such a diverse set of objectives was cumbersome and created additional challenges. The university needed an immediate and scalable enterprise platform to manage data and analytics, and it needed to be accessible while maintaining privacy and security standards.

Solution
With funding from MSSU, HDNS embarked on upgrading their infrastructure to move off on site servers to a local-cloud solution, turning to Cloudera's platform for data management. Delivering security and remote access was a top priority as this enabled them to address both business and technical needs. Selecting Cloudera Data Platform (CDP) and Cloudera Professional Services to drive a big data strategy forward became the logical choice.

"The vision was that we didn't just want a system that could support what we do, but we wanted one which could support what people might want to do in the future. That was the reason to make the investment in Cloudera," said Dr. Samuel Stewart, Director, Health Data Nova Scotia (HDNS).

> "The vision was that we didn't just want a system that could support what we do, but we wanted one which could support what people might want to do in the future. That was the reason to make the investment in Cloudera."
>
> Dr. Samuel Stewart, Director, Health Data Nova Scotia (HDNS)

Cloudera Professional Services shortened time to value by using tried and tested migration methodology, automation tools and deep expertise. The result is a future proof target state architecture that will help HDNS drive exponential business value throughout their data journey.

Delivering secure patient-oriented analysis

Dalhousie provides a platform for both HDNS staff and external analysts to access data for research projects. Their future data holdings beyond health information will include social information, education, justice and other government data, along with large research-specific datasets, offering a platform for generating new health insights.

Security is of utmost importance and must support all data forms. In healthcare, patient data with names and addresses go through the same level of security as de-identified data that has no names and addresses and has encrypted IDs.

"In Nova Scotia, the province where Dalhousie University is located, the Personal Health Information Act (PHIA) is the governing body, but it's different in each province. All the provincial data centres are individually regulated by their province, and with public health legislation. It's very provincial in Canada, we don't have anything quite like the United States Health Insurance Portability and Accountability Act (HIPAA) from a federal perspective," said Stewart.

Supporting the business and infrastructure

HDNS was looking to support both the business and technical infrastructure. "We saw value in the ease of use managing the platform with additional hard drives, three-way replication of all files, and we did not need to back up to external sites. In a case where a hard drive fails we are easily able to add a new one - the platform builds itself," said Jordan Farrell, System Administrator at HDNS.

With the planned upgrade to CDP Private Cloud, HDNS will have reliability and scalability, as well as speed of processing on a distributed cluster, and support for statistical programs such as SAS. From Stewart's perspective, "what we are doing with the remote access system wouldn't have been possible with our old structure—the backend database wouldn't have been able to support this many concurrent connections. This system has powerful security that can support all forms of data."

The university's external analysts access data through SAS Enterprise Guide where the project libraries are linked directly back to the core data. "These are actually databases, built into the platform, and analysts have no idea they're not even touching SAS files. Secretly a huge advantage of this system is it allows multiple users to be using the SAS interface, while interfacing with the Cloudera platform," added Stewart.

Driving machine learning innovation on new projects

Machine learning (ML) is important to the university because they support any research people are interested in doing. Today, HDNS supports Python and R for machine learning, and they are open to supporting new projects that will benefit everyone. The improved data infrastructure has allowed HDNS to support ML more completely, which has expanded the user base overall. They are delivering more projects and inquiries, and that will continue to increase as the health research community embraces ML more completely.

Results

With the enterprise platform in place, HDNS' big data goal to drive insights more quickly is being achieved. They now support external researchers, with exponential growth for remote users.

"Our external analysts in epidemiology or genomics don't see the raw data; they can only see what the research question allows them to see. Previously, they had to go into a room and work on a small Windows server to do their analysis." Farrell added, "This was a manual process for them to schedule a few hours to get the work completed. Now remote workers are able to get data served up faster and with better reliability, and all in a cloud-native manner."

CUSTOMER SUCCESS STORY

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

Engaging Cloudera Professional Services and assessing the scope of the migration was a simple process. They quickly gained an understanding of the migration and future usage of CDP and constructed a migration plan that was tailored to their requirements. Having a highly skilled Solutions Architect during the migration will allow the Dalhousie team to make the move to CDP quickly and have the platform configured in a way which allowed them to leverage the value right away.

Dalhousie University will be working with the Canadian Health Data Research Network (HDRN) in distributed analysis in which individual data-holding centres work together to create a single portal and support system for researchers requesting multi-jurisdictional data. The goal is to develop technological infrastructure to improve data access and collection, create support for advanced analytics and establish strong partnerships with patients and the public. "It's a centralized analysis where the data is housed in different locations and analyzed as if it's a single data holding. The data is held in multiple provinces and never moves between centres, it's the analytic components that will be moved between centres," said Stewart.

HDNS' vision moving forward is that everything will continue to be accessed this way, and in the private cloud. This secure and locked down remote connection allows them to take on more projects and expand their user base. "I feel like we have more projects every year than the year before, and more distributed projects. The advantage is that we're now prepped for the future as things go forward and we expand our data holdings," added Stewart.

CLOUDERA

Cloudera, Inc. 5470 Great America Pkwy, Santa Clara, CA 95054 USA. cloudera.com

© 2021 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice.