CASE STUDY

Quebec Blue Cross Achieves Full CI/CD for Data Analytics with Dremio's Lakehouse Platform

At a Glance

The Customer

Challenge

Quebec Blue Cross's small six-person analytics team struggled with manual SQL deployment processes that were error-prone and time-consuming. Their previous cut-and-paste deployment method created deployment nightmares, taking hours to complete and causing significant stress for the team while limiting their ability to scale data operations effectively.

Solution

Quebec Blue Cross implemented Dremio as their SQL engine on top of their Azure data lake, integrating it with DBT Core and Apache Airflow to create a fully automated CI/CD pipeline. This infrastructure-as-code approach enabled automated deployment of virtual datasets, comprehensive testing, and streamlined data operations management.

Results

Transformed deployment from manual 3-hour processes to automated 4-minute deployments, scaled from 200 to 1,200 physical datasets and 950 to 1,400 virtual datasets within 12 months, while enabling the small team to deploy dozens of validated objects weekly with dramatically improved quality and user satisfaction.

The Customer

Quebec Blue Cross operates as a nonprofit travel insurance organization dedicated to providing comprehensive services to Canadian travelers worldwide. The company literally saves lives through their emergency medical services, handling everything from medical evacuations to mental health support for travelers in crisis situations. As a mission-critical organization that coordinates complex international medical emergencies and travel assistance, Quebec Blue Cross relies heavily on data analytics to optimize their operations, monitor service quality, and make rapid decisions that can impact customer safety and well-being across global destinations.

The Challenge

Quebec Blue Cross faced significant operational challenges with their data analytics infrastructure despite having assembled a dedicated six-person analytics team plus DevOps support. Their primary struggle centered around manual deployment processes that had become increasingly unsustainable as their data operations grew. The team was using a cut-and-paste method for deploying SQL statements and virtual datasets, which created what their architect described as "deployment nightmares" that were driving the team to distraction.

These manual processes were not only time-consuming, often taking three hours to complete deployments, but also highly error-prone and impossible to scale effectively. With the organization's growing reliance on data-driven decision making for their life-saving services, the team needed to dramatically increase both the quality and speed of their data product deployments. The manual approach severely limited their ability to respond quickly to new use cases, planned projects, strategic initiatives, and the unplanned emergency situations that frequently arise in their travel insurance business.

Additionally, as a small team managing an increasingly complex data ecosystem spanning Oracle databases, REST APIs, SharePoint, Salesforce, and various other data sources, they needed a solution that would enable them to maintain and evolve their infrastructure without overwhelming their limited resources. The lack of automated testing and validation meant that deployment quality was inconsistent, directly impacting the user experience for stakeholders who depended on accurate, timely data for critical business decisions.

The Solution

Quebec Blue Cross implemented Dremio as the cornerstone of their modern data architecture, positioning it as a massively parallel SQL server on top of their Azure data lake. This implementation was part of a carefully orchestrated solution that leveraged each tool for its core strengths while maintaining simplicity for their small team to manage effectively.

The architecture integrated Dremio with DBT Core to enable infrastructure-as-code practices for their data transformations and virtual dataset creation. Using Apache Airflow as their orchestration engine, they built a comprehensive CI/CD pipeline that automated the entire deployment process from development through production. Their approach utilized Dremio's virtual dataset capabilities to provide a semantic layer on top of their bronze, silver, and gold data lake layers processed through Databricks.

The team developed a streamlined workflow where analysts and engineers develop and test SQL statements directly in Dremio's interface, then transform them into DBT syntax for version control and automated deployment. Their CI/CD pipeline includes automated backup procedures, synchronization with Bitbucket for source control, automated group and user synchronization with Azure Active Directory, and comprehensive validation of both physical and virtual datasets through automated testing.

A key innovation was their implementation of automated validation processes that perform select queries against every object in Dremio to ensure both physical datasets and virtual datasets remain valid after deployment. They also added metadata tags to distinguish between DBT-managed objects and user-generated content, enabling better change management and object lifecycle control.

Results

The transformation delivered remarkable improvements across all aspects of Quebec Blue Cross's data operations. Most dramatically, deployment time decreased from three hours to just four minutes and seven seconds, representing a 98% reduction in deployment duration while simultaneously handling significantly more complex deployments with thousands of objects rather than the hundreds they previously managed.

The team achieved unprecedented scalability, growing from 200 to 1,200 physical datasets and expanding virtual datasets from 950 to 1,400 within a single 12-month period. This six-fold increase in data assets was managed by the same small team, demonstrating the power of their automated approach. They now deploy dozens of fully validated and tested objects every week, a volume that would have been impossible under their previous manual processes.

Quality improvements were equally significant, with deployment success rates exceeding 95% through their automated validation processes. The consistent, reliable deployments eliminated the stress and errors associated with manual processes, leading to much happier users who now receive validated, high-quality data products. The analytics team reported dramatically improved job satisfaction due to the elimination of tedious manual operations and the ability to focus on higher-value analytical work.

The infrastructure-as-code approach enabled true collaborative development, where team members can develop independently, test thoroughly in development environments, and deploy through standard pull request workflows. This professional software development approach elevated their data operations maturity while providing the audit trails and change management capabilities essential for their regulated industry environment. The automation freed up significant time for the team to focus on new use cases and strategic initiatives rather than maintenance and deployment activities.