Case Study

Training and Education Organization Learns How to Get More from their Snowflake Data Warehouse Analytics Environment

by Tobin Thankachen

aws PARTNER

- Public Sector
- Immersion Day
- Amazon RDS Delivery
- Amazon EC2 for Windows Server Delivery
- AWS Microsoft Workloads Services Competency

datAvail

BI/Analytics • Applications • Databases

Introduction

With enterprise technology constantly evolving, even the cutting-edge tools and best practices of yesterday can become outdated tomorrow. For this reason, it's an excellent idea to schedule thorough health checks and assessments from a trusted IT partner at regular intervals.

In this case study, we'll discuss one of our recent clients, a nonprofit in the education industry, which was interested in refreshing and strengthening its IT ecosystem. Datavail provided a number of recommendations across the board, from data integration to software development, to help the client improve its reporting and analytics workflow.

The Client: Education Nonprofit with Data-Driven Needs

The client in this case study is a U.S. nonprofit job training organization that has helped thousands of young adults prepare for careers in information technology and financial services. With dozens of campuses across the country and corporate partners that include many Fortune 500 firms, the client is devoted to helping talented, economically disadvantaged students develop their professional skills.

Prior to working with Datavail, the client was pulling data from a number of different sources. These included CRM and ERP software such as Salesforce, NetSuite, and TargetX, the financial planning and forecasting software Adaptive Insights, and the labor market analytics tool Burning Glass.

Specifically, the client used the ETL (extract, transform, load) software Fivetran to capture the latest changes to their engagement data and then load these changes into a Snowflake data warehouse. In addition, the client was using the Matillion ETL platform hosted in Amazon Web Services EC2 to ingest data from all of their sources, transform this data, and load it into Snowflake. Once this information was inside Snowflake, the client was also using the open-source tool dbt to transform data on the fly for reporting and analytics.

The client was interested in getting an assessment of their current data reporting and analytics environment, as well as identifying any performance or architectural issues. The client's specific goals for the assessment included:

- Determining whether the client was optimally applying tools and technologies within their IT architecture.
- Finding ways to improve performance and remove existing bottlenecks.
- Increasing the use of automation within the client's current data ingestion processes.
- Suggesting necessary changes to support higher data or transaction volumes in order to improve scalability.
- Fine-tuning processes to ensure that the appropriate audits, alerts, and controls are in place.

The Problem: Issues with Data Quality, Security, Performance, and More

There were several issues with the client's existing data reporting and analytics workflow. The client was struggling with the complexity of their IT ecosystem; as a small nonprofit organization, the client also lacked the in-house staff necessary to handle the challenges of their data reporting and analytics environment.

Data integration

As mentioned, the client was using multiple tools for extracting and loading data, including Fivetran and Matillion. This convoluted system was due to accumulating software over time, with new tools added as their IT team grew.

However, the choice to use several data integration tools added greater overhead and cost to the client's data integration pipelines. In addition, the client was not entirely certain of which data sources they were using, with several of these sources obsolete (i.e., no longer providing relevant information).

Snowflake

The client's issues continued within the Snowflake data warehouse. Previously, the client had been using Microsoft SQL Server, which was not enough for their needs, before migrating to Snowflake. The client was now unsure whether they were getting the most from their use of Snowflake.

Problems with the client's Snowflake deployment included:

Lack of separation

The client was using a single database to hold the transformed data, rather than separating data across development, test, and production environments.

Backups and business continuity

The client did not have a mature backup strategy for Snowflake data.

Data quality

The client was not running tests to validate the quality of incoming raw data, or of the staging data generated after cleansing the raw data.

Data privacy

The client was not adequately performing data masking to conceal the contents of sensitive and confidential information.

Data security

The client allowed users from any IP address to access Snowflake, posing a significant security risk.

Software development

Other issues surrounded the client's software development workflow. Although the client was trying to implement a CI/CD (continuous integration/continuous delivery) and DevOps pipeline, there were several barriers holding the client back from optimal efficiency and productivity.

The problems with the client's CI/CD and DevOps pipeline included:

Access control

All developers had access to the production environment, and all developers could merge code to production in GitHub without requiring approval. The client used only a single repository for both the development and production environments.

Missing QA environment

The client had separate environments for development and production, but not for quality assurance (QA). Introducing a third QA environment would enable the client to improve data reliability with a more holistic testing suite.

Lack of automation

The client's CI/CD process required a good deal of manual effort. For example, after creating a pull request, developers had to manually merge the development branch with the master branch and delete the existing dev branch. The client was not taking advantage of automated processes that could reduce developers' effort.

Tableau

Last but not least, the client was also suffering a number of pain points with its Tableau data visualization software. The concerns here included:

- A lack of concrete standards and best practices (e.g., regarding file permissions and distributing insights throughout the organization).
- Performance issues with multiple views and dashboards.
- Roughly 10 gigabytes of unused data.

The Solution: Robust Assessment of Reporting and Analytics Workflow

Facing a number of issues and challenges in its IT environment, the client decided to work with a skilled, experienced data reporting and analytics partner. Specifically, the client wanted to find the right managed services provider who could help diagnose their problems and propose the right improvements and solutions.

In search of the best company to join forces with, the client reached out to Datavail. One of the client's IT directors had previously worked with Datavail at another organization, making use of Datavail's enterprise database services.

Thanks to this positive experience, the director convinced the client's chief information officer that Datavail had the expertise, skill set, and resources to make the assessment a success. The client was also pleased that Datavail had an in-house team of full-time employees to perform the work, rather than hiring third-party consultants.

After discussing the terms of the relationship, Datavail sat down with the client's staff and began to analyze their IT environment. This included everything from the client's tools and job flows to the underlying data structures and final analytics reports.

Recommendations

Datavail's main conclusions were as follows:

- The client was using multiple toolsets and vendors for extracting and loading data, resulting in overlapping and redundant functionality.
- The client's current reporting and analytics environment was scalable—an important concern since the client was planning for future growth.
- However, the processes and controls surrounding the client's production data and code library were too informal.
- The client was performing too little data validation and masking to ensure a high level of data quality and security.
- In general, the client had several areas of improvement for its IT and data security policies and procedures.
- Despite these issues, the client enjoyed a reasonably good level of reporting and data visualization.

Given these findings, Datavail offered the following high-level recommendations:

Data integration

The client should reevaluate their use of existing tools for data extraction and loading. In particular, the client should either justify and rationalize the current setup or move to products and vendors with a lower cost and SaaS (software as a service) model. Datavail recommended that the client continue their use of Fivetran while adding a new data integration tool, AWS Glue.

Software development

The client should introduce several improvements in order to further embrace CI/CD and DevOps. In particular, the client should create three distinct development, test, and production environments, as well as automate and formalize the processes surrounding code promotion and approval. Datavail also recommended upgrading the dbt data transformation tool from dbt Core to dbt Cloud.

Feature adoption

The client should make use of existing features and functionality in their IT environment to improve its data integration and software development workflows. For example, the client should use Snowflake's zero-copy cloning feature to replicate data from the production environment for testing. In addition, the client should better protect their Tableau assets by using folders to better emulate separate development, test, and production environments.

Data quality

The client should increase their data quality through greater adoption of the dbt standard, as well as using custom data validation features. Also, the client should go through all uses of primary data sources to determine if they are still providing useful information.

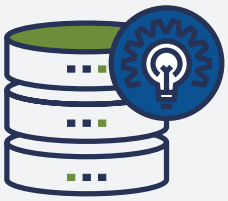
Data security

The client should boost their data security practices by enacting additional network security policies and using the data security features of Snowflake and Tableau. For example, the client should increase the use of data masking in Snowflake to protect sensitive and confidential information.

Concretely, Datavail made the following recommendations for software upgrades and changes:

Data extraction and loading

The client was using a combination of Fivetran, Matillion, and Python scripts. Datavail recommended the client switch to a combination of Fivetran and AWS Glue.



Data transformation

Datavail recommended the client upgrade its dbt tool from dbt Core to dbt Cloud, a more robust integrated development environment (IDE) with advanced development features.

Version control

Datavail recommended the client migrate from GitHub Free to the paid version, GitHub Enterprise, in order to access additional functionality for CI/CD, DevOps, and automation.

Snowflake

The client was using a single database for all environments. Datavail recommended the client use multiple databases for development, QA, and production, and use Snowflake's zero-copy cloning to replicate this data.

Data extraction and loading

First, Datavail compared the client's existing setup for data extraction and loading with other leading platforms. These comparisons were based on factors important to the client, such as the cost of conversion, support, and infrastructure.

Datavail's in-depth comparison analyzed all the options along a wide range of axes. For example, Datavail found that Stitch and Fivetran were the best data integration tools for low-volume workloads, while Fivetran would be costlier than Stitch if workloads primarily consisted of inserts. Meanwhile, AWS Glue and Apache Airflow had the lowest infrastructure cost, although Airflow had less data security relative to other alternatives.

In total, Datavail recommended that the client convert 40 data extraction and loading pipelines to Fivetran and AWS Glue. By using the proposed ELT architecture and tools, the client would make its data integration processes simpler and more scalable.

Software development

Next, Datavail recommended that the client increase their use of automation for their CI/CD workflows. When a pull request occurs, a continuous integration job runs to ensure that the new code can be safely merged into the existing repository. If the job fails on any validation checks run against a cloned production instance, a notification is automatically sent to the workflow manager or lead developer. This individual can then manually verify these failures and discuss how to fix the problematic code with the responsible developer.

This setup has a variety of benefits. Most importantly, it maintains the security of production code by ensuring that all code goes through a series of validation and approval steps. This helps the client improve its adoption of DevOps in their production environment.

Datavail also recommended that the client migrate from the free version of GitHub to GitHub Enterprise. The paid version of GitHub would provide benefits such as enhanced security (in the form of single sign-on and IP whitelisting) and multiple pull request viewers.

Data security

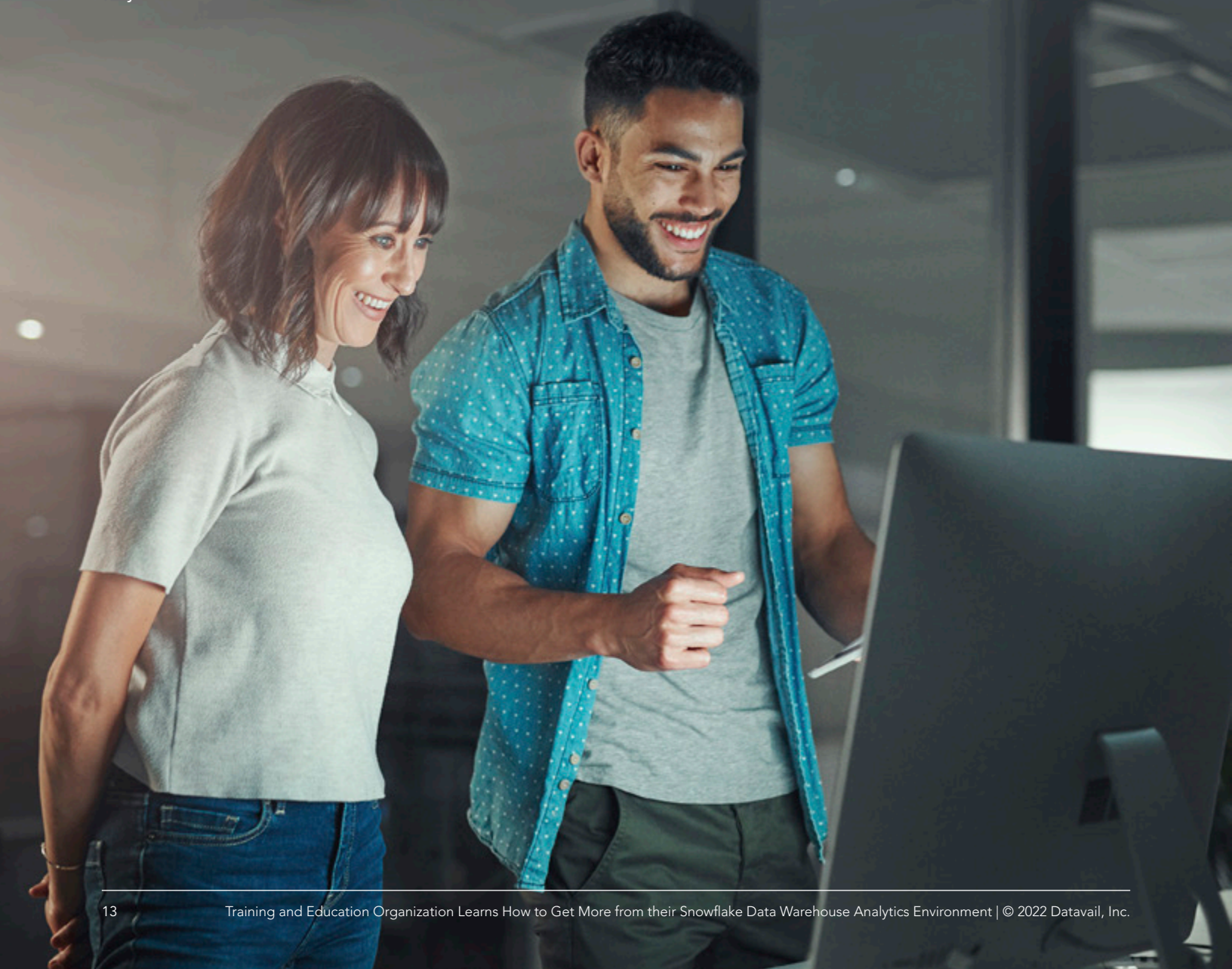
Datavail offered a number of recommendations across the client's IT environment to improve data security. These included:

- Implementing a robust **network policy** by limiting the IP addresses from which users can connect to Snowflake, stopping many attackers in their tracks.
- Creating a strategy for **data backups and business continuity**, helping the client avoid disruption in the event of a data breach, ransomware, or other cybersecurity event.
- Deploying **role-based access control (RBAC)**, decreasing the likelihood of a breach by controlling users' access to data in different environments.
- Using **data masking** to hide sensitive and confidential information while in transit, preventing unauthorized people from viewing it.
- Enabling **multi-factor authentication (MFA)** for all users, helping to secure data even if passwords are compromised.

The Results: A Clear Roadmap for the Future

The client now has Datavail's extensive list of recommendations, cost-benefit analyses, comparisons, and more at their fingertips. Thanks to Datavail's comprehensive IT assessment, the client has a crystal-clear roadmap for how to improve their data reporting and analytics.

For now, the client plans to implement these recommendations progressively in-house, using new team members hired since the departure of former key personnel. In particular, given the client's future intent to scale their data consumption and analytics, the client sees Datavail's recommendations surrounding data integration tools and Snowflake data warehousing as invaluable. Of course, the door remains open for Datavail to work closely with the client on another IT project, as best fits their business needs and objectives.



How Datavail Can Help

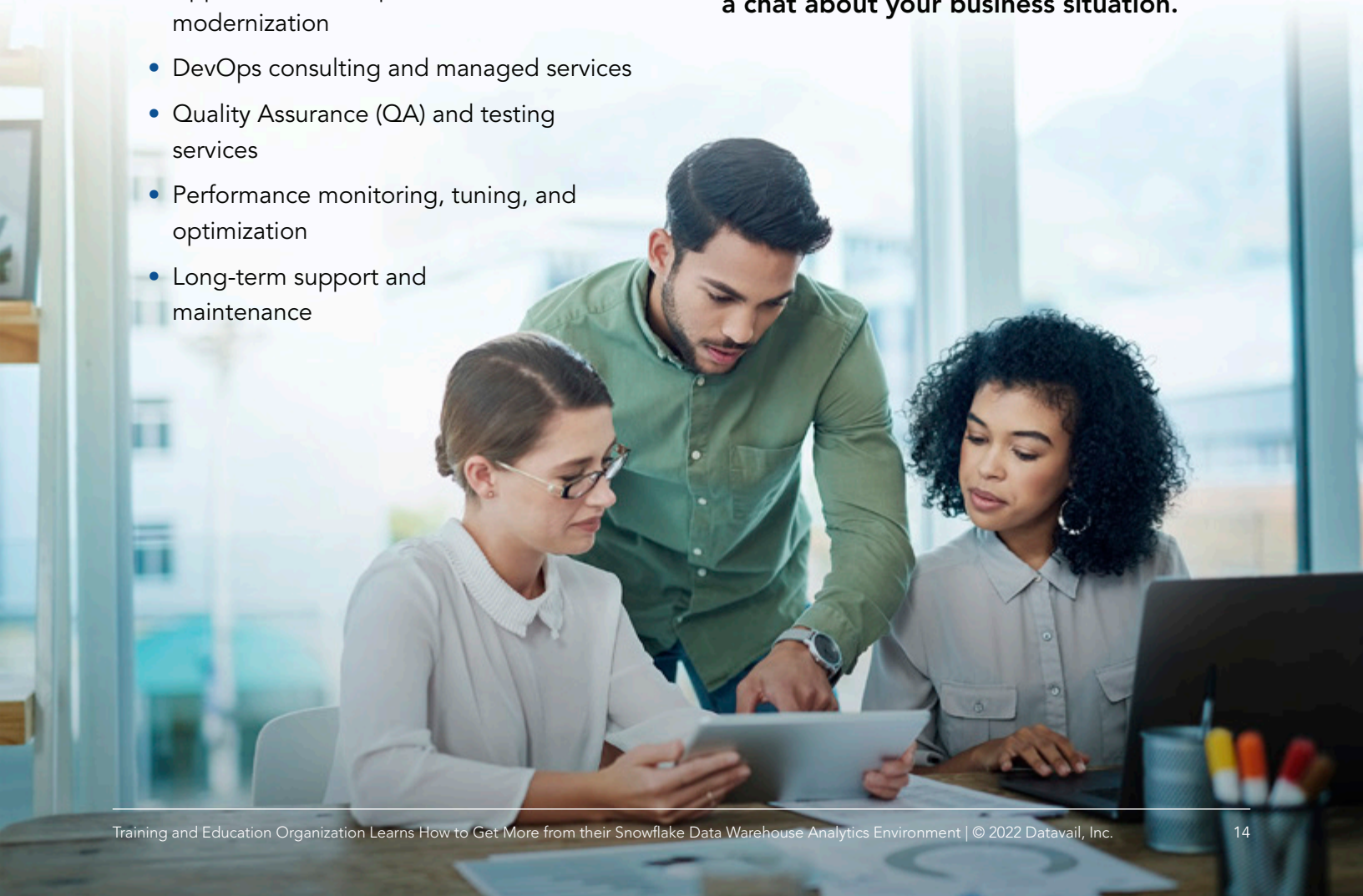
Datavail's work in this case study is just one example of how we help companies turn their enterprise data into a strategic driver of their business. We are a knowledgeable, qualified, and highly experienced IT partner for thousands of clients who need assistance with their analytics, business intelligence, databases, and application development.

Datavail offers an end-to-end suite of business intelligence and analytics consulting services, including:

- Health checks and assessments
- Strategic planning and roadmapping
- Data integration and warehousing
- Application development and modernization
- DevOps consulting and managed services
- Quality Assurance (QA) and testing services
- Performance monitoring, tuning, and optimization
- Long-term support and maintenance

With countless satisfied clients and an average partnership length of more than 7 years, Datavail has what you need to make your next IT project a success. Datavail has acquired an extensive list of strategic partner alliances, including an AWS Advanced Tier Consulting Partner for applications, databases and analytics; an Oracle Specialized Partner for business intelligence and databases; and a Microsoft Gold Partner with experience across the Microsoft technology stack.

Ready to learn how Datavail's health checks and assessments can help transform your IT environment? Get in touch with our team of experts today for a chat about your business situation.



Biography

Tobin Thankachen

Lead Architect, Analytics

Tobin Thankachen is Lead Architect at Datavail, and a proficient Cloud & Data Analytics Lead with strong leadership and solutions expertise in Cloud, Big Data and Traditional Data warehouse. He has developed strategies for accommodating modern use cases for data delivery such as large data volume, unstructured data, data discovery, cognitive and data science analytics.

Tobin has also spear-headed organizational objectives by leveraging Cloud Data Migration, completing performance tuning, assessments, roadmaps and recommendations in the Analytics space. Additionally, he has also led cross-functional projects using advanced data modeling and analysis techniques to discover insights that will guide strategic decisions and uncover optimization opportunities.

Improving organizational performance, Tobin evaluates best practices for DB servers and data quality issues for ETL and Analytics systems.

datAvail

BI/Analytics • Applications • Databases

About Datavail

Datavail is a company of over 1,000 professionals helping clients build and manage applications and data via a world-class tech-enabled delivery platform and software solutions across all leading technologies.

visit us

Corporate Headquarters
Datavail Corporation
11800 Ridge Parkway, Suite 125,
Broomfield, CO 80021

Give us a call
877-634-9222

Email us
info@datavail.com

www.datavail.com