Case Study

How a Data Expert Improved its Data Integration, Storage and Reporting

by Tobin Thankachen

datAvail

BI/Analytics • Applications • Databases

Introduction

Sometimes, even data management professionals need help with data management.

This case study examines Datavail's strategy to modernize the data capture, conversion, storage, management, and retrieval system structure used by a data collection and aggregation services company. Their original infrastructure was, in its day, very cutting-edge, and provided their customers with top-of-the-line data research and retrieval capacities. Technological advances, however, had overtaken even their sophisticated architecture, and they could no longer provide the fast, comprehensive, and efficient reporting that their patron base had come to expect.

They asked Datavail to evaluate their system for challenges, offer insights and suggestions for improving it, and, ultimately, perform the services necessary to bring it to current, modernized, cloud-based standards.

The Situation

The Client provides its customers access to a series of comprehensive business- and business management-related databases. Its programming scours corporate, government, regulatory, and other state and federally based resources for information on various organizational intelligence and activities. It designed its 14 distinct 'products' to address specific customer inquiries, from diversity trends to international corporate information. Customers search through these databases to find information about individual companies, corporate leadership, relevant regulatory actions, etc. Culled from global data sources, the aggregate database provides a comprehensive collection of business activities around the country and continent. Users can seek and find high-quality, accurate, and timely information about industry regulations, corporate names and registrations, corporate title transfers, business registries, and more.

The challenges presented by the Client's existing on-prem configuration were many and varied:

- Because its scope encompasses the continent, its central processing servers received data in disparate forms and types from all those far-flung agencies.
- Complicating the capture function were the file patterns, table structures, and package rules that differed from state to state and country to country.
- The volume and complicated nature of the overarching data flow created bottlenecks and numerous errors and prevented sequential execution of package procedures.
- The original architecture required 14 separate steps to capture, extract, transform, and load (ETL) product information into its final location on the central corporate database, where users could find it.

Not surprisingly, the Client and its customers experienced frequent challenges when requested data arrived with errors, delays, or didn't arrive at all.

In addition to these technological challenges, the Client was also struggling with its database management capacities in general. They had recently experienced significant staff turnover and had lost much of their database management (DBA) staff. There were few data technicians still around who understood the configuration of their on-prem resources, and they had just one in-house developer. This employee was tasked with handling both the systems development management and the reporting functions.

The situation was complicated:

- They were very concerned about data security and worried they could not defend against emerging global cyber security threats.
- They were also subject to HIPAA regulations and were not at all confident that their systems could pass those standards.
- The loss of staff was extremely concerning. Replacing knowledgeable technology professionals was almost impossible since their original infrastructure was built on fading legacy tech.

Despite providing a unique and valuable service to the global business community, this enterprise struggled with its capacity to maintain its high-quality databases, products, and services.

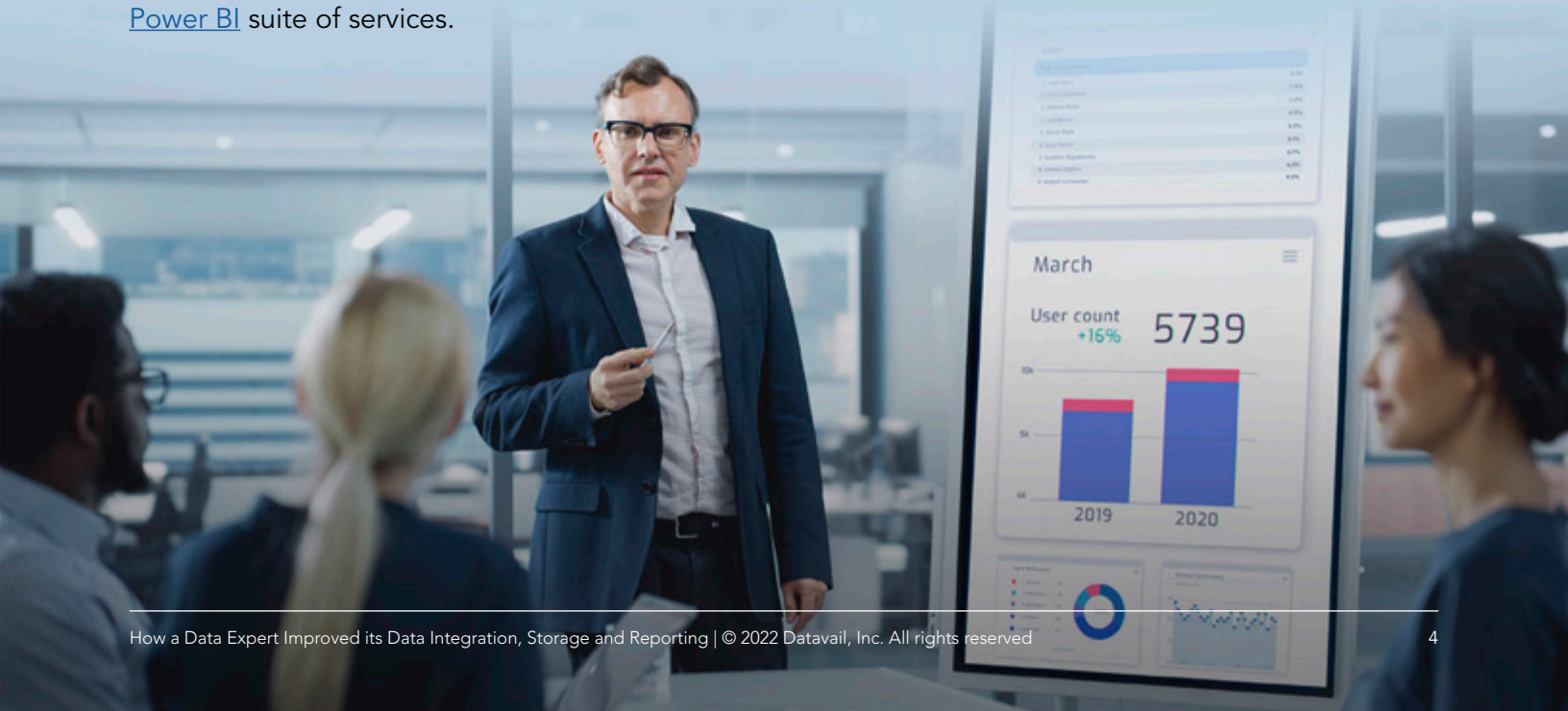
The Assessment

Datavail's approach to the challenge began with assessing the situation to determine which services and products it could provide to address the constellation of Client concerns.

Initially, the Client was inquiring about database management assistance. Their lack of technical staff was having a negative impact on their corporate activities, and they understood that Datavail offered exceptional database administration services. So Datavail had already begun providing this Client with SQL database support when its leadership came asking about the tech company's development and analytics capacities. Over the course of a few weeks, the two organizations collaborated on what the Client's infrastructure required to achieve its corporate goals and the cloud-based products and services that Datavail could provide to accomplish those gains. Datavail's cloud computing professionals then evaluated the existing on-premises architecture and developed a roadmap to migrate application functions from that on-prem location to the Azure Data Factory (ADF) and the Microsoft Power BI suite of services.

The complexity of the project suggested that Datavail present it as a two-part project, the first part explaining the concept behind Datavail's proposal - a 'Proof of Concept' (PoC). A PoC is itself a project as the contractor develops and defines the resources they would use to address Client concerns and meet Client requirements. In this case, the Client was uncertain about the sufficiency and capacity of cloud computing assets as compared to their trusted, on-prem legacy technology suite.

The second part presented the value provided to the Client of the re-envisioned framework -- a Proof of Value (PoV). A PoV ties the effort of the PoC to the anticipated business value of the proposed solution. For any company, it's not enough to understand how technology will improve or enhance its performance. These days, that technology must also encompass the fundamental corporate purpose and be designed to achieve its specified goals. Datavail's proposal needed to provide technical solutions to performance problems while also offering growth opportunities.



The Proof of Concept

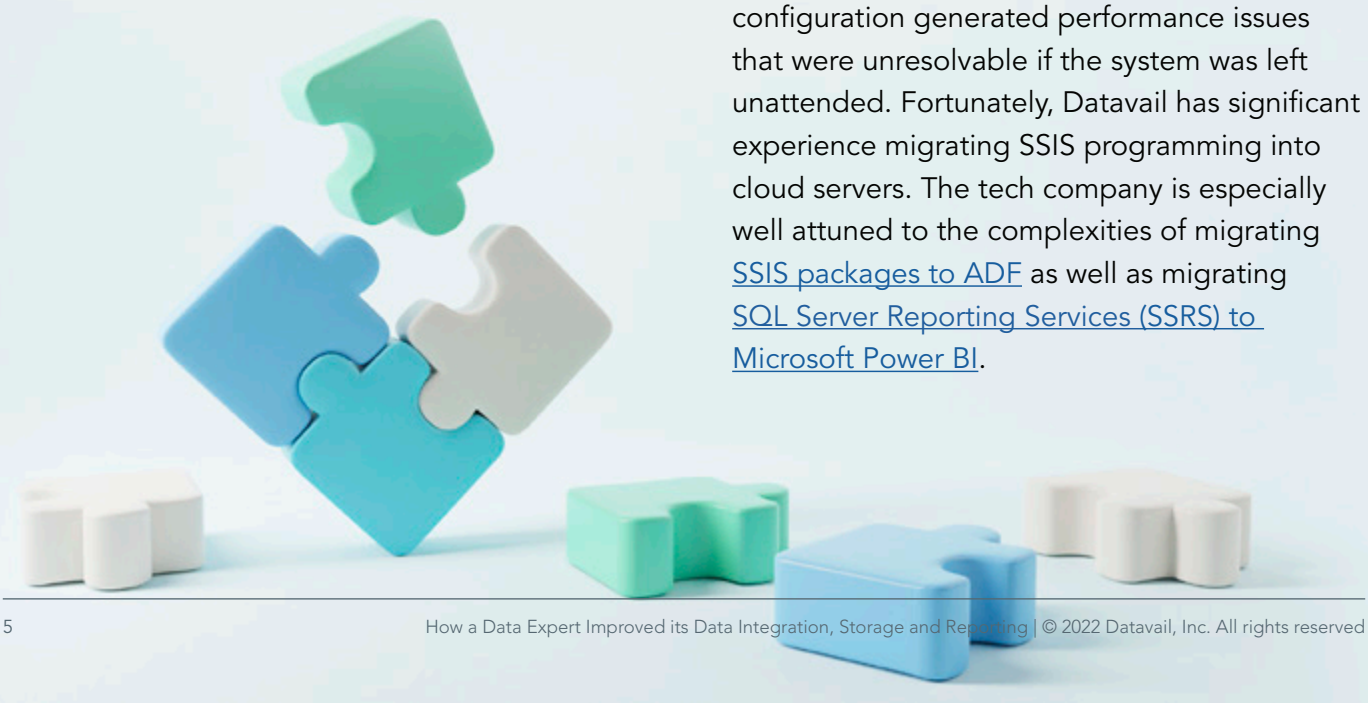
There was no way for Datavail to find and assess all the issues arising across all 14 products. Instead, the data management experts elected to analyze one unique product to determine where bottlenecks, errors, and other interferences were occurring within its processes. Presumably, errors found there would be repeated throughout the other services, too.

Fundamentally, the legacy architecture used SSIS packages (SQL Server Integration Services) as its data capture, Extract-Transform-Load (ETL) functions, and repository tools. An SSIS package assembles a series of connections, data flow elements, control flow elements, and event handlers into a single unit with the capacity to execute a particular ETL task. The integration application includes a proprietary set of built-in tasks, graphical tools to structure the package, and a catalog of databases for storing, running, and managing packages. It can extract and transform information from numerous data sources, including XML data files, relational data sources, and flat files. In their day, SSIS packages provided exceptional ETL resources for on-prem activities.

The SSIS-based system that Datavail discovered was rife with errors, breakages, and gaps. The company used one SSIS package per state, and each package was hardcoded for the process specific to that state. Consequently, packages streaming from each state comprised their own individual and unique coding, parameters, instructions, etc. Errors, bottlenecks, and gaps required manual attention and interventions, and those skill sets had to be tailored to the coding, etc., unique to each state. Making matters worse, the enterprise used several separate databases, which also required many manual inputs throughout the ETL process.

Additionally, the hardware architecture was also complicating things while negatively impacting system performance. The Client was running both its SSIS package system and all its databases on one server. As a result, usage volume, data flow, and user access demand all hampered the performance of the system's already complicated infrastructure.

Together, the technology and physical configuration generated performance issues that were unresolvable if the system was left unattended. Fortunately, Datavail has significant experience migrating SSIS programming into cloud servers. The tech company is especially well attuned to the complexities of migrating SSIS packages to ADF as well as migrating SQL Server Reporting Services (SSRS) to Microsoft Power BI.



The Solution

To identify the specific problems and possible solutions, Datavail analyzed the inner workings of the Client's product for use as a testing opportunity and solution example. This information laid the foundation for the Proof of Concept (POC):

- A thorough assessment of each process step was required learn how data moved through the ETL process and into a fully integrated database.
- The Datavail team sorted through individual functions, data flows, workflows, parameters, etc., based on the company's actual activities to learn how they integrated the information per state and then integrated it again within the Clients' database as a whole.
- They then diagramed the data flow to create a visual representation that explained to the Client where within its systems:
    - challenges were arising,
    - performance was declining, and
    - how modernized cloud-based programming could provide enhanced performance capacities while also eliminating all these technical challenges.

Datavail's proposed solution was exceptionally simple when compared to the Client's original setup:

- They suggested moving data contained in on-prem File Transfer Protocol (FTP) and Secure File Transfer Protocol (SFTP) sources into an Azure Data Factory data lake (ADF). The ADF programming eliminated the use of the SSIS packages, which also eliminated the challenges those posed.
- In ADF, while data flowing from each state remained segregated per state, the ADF cloud reconfigured all those individual sources and directed them into a single, state-dedicated pipeline. Within each channel were all the technological functions required to collect, cleanse, and integrate the information for loading into the Client's corporate databases.
- In addition to recommending automated ADF processing of ETL file transfers within each state's pipeline, Datavail also recommended automating their subsequent loading process into a 'parent' pipeline that would integrate the data from all states. This 'parent' pipeline would then feed the information into the file formats used in corporate database tables. Through both processes, automation capacity eliminated the need for expensive human oversight, and cloud-based applications could simultaneously provide the automation function and ensure its accuracy.
- Once the new database was assembled and populated, Datavail recommended funneling corporate information through Microsoft's Power BI program for reports and analysis.

Proof of Value

The Client asked Datavail to develop a PoV around the use of ADF and Power BI for this project. Their concern was determining the budget parameters of the migration of their databases and programming from the on-prem configuration to a cloud-based structure. Datavail's documentation of its SSIS to ADF migration concept demonstrated how one-half of the project could be accomplished. Applying the Microsoft BI programming gives cloud engineers the strategic, security, and growth software they need to architect the new cloud-based system to address and encompass the Client's business goals.

The Solution

In addition to enhancing performance capacities and automating most of the underlying data and database management infrastructure, the BI suite of services overlaid additional but required structure and parameters to ensure that the Client's investment delivered an appropriate return.

They specialized configuration by state

Just as the Client's business was complex, so would be the resulting cloud configuration of its processes. As noted above, each state pipeline contained a multitude of varied information streams flowing from a variety of sources. Once those data were integrated into the Client's central database, deriving intelligence and insights from their combined mass required configuring reports based on the specific state's criteria, as well as those needed by the Client's customers. Datavail proposed using BI tools and technologies to extract and reveal the observations contained within those corporate files and designed report formats and templates to address the needs of individual users. The systems built into each state platform were then analyzed to provide comprehensive business intelligence oversight.

They were accurate

Also imperative to the project's success was the accuracy of the finally populated database. The data flowing from each state resource required "cleansing" before its initial integration with the other states' information to ensure that it was truthful, complete, and relevant. The same cleansing process was required as state databases were integrated into the central corporate database and again as those fed relevant data into the various and appropriate subject matter report forms and formats. Configuration of programming assets at this stage added audits, controls, and alerts to ensure that final reports were based on accurate, appropriate, and timely information.

They were secure

Many C-Suite residents only think of "cybersecurity" when considering the security needs of their digital infrastructure, and they're not wrong. However, while cybersecurity - protecting against external intrusions and breaches - is definitely required, internal security protections are also mandated by both stakeholders and industry regulators. To ensure only appropriate access is available to users, the BI suite of services maintains a vigil over unique users, preventing those without proper authority from gaining entry to sensitive corporate intelligence. Extending this capacity includes constant assessment of shifting access privileges, rolls, and requirements, as the Client grows or contracts its Clientele, workforce, contractors, and other entities needing access to its database.

They were scalable

Optimally, the reimagined data store would also have the capacity to grow as the Client added new options, services, or assets. Datavail engineered the ADF/BI architecture with an eye on the future so that the Client need only request an expansion of its assets and not a complete rebuild of its configuration. Without question, the capacity for scalability was not available on the Client's former on-prem system.

A review of the resulting proposal revealed that it encompassed what many believe to be "best practices" for cloud computing configuration and use:

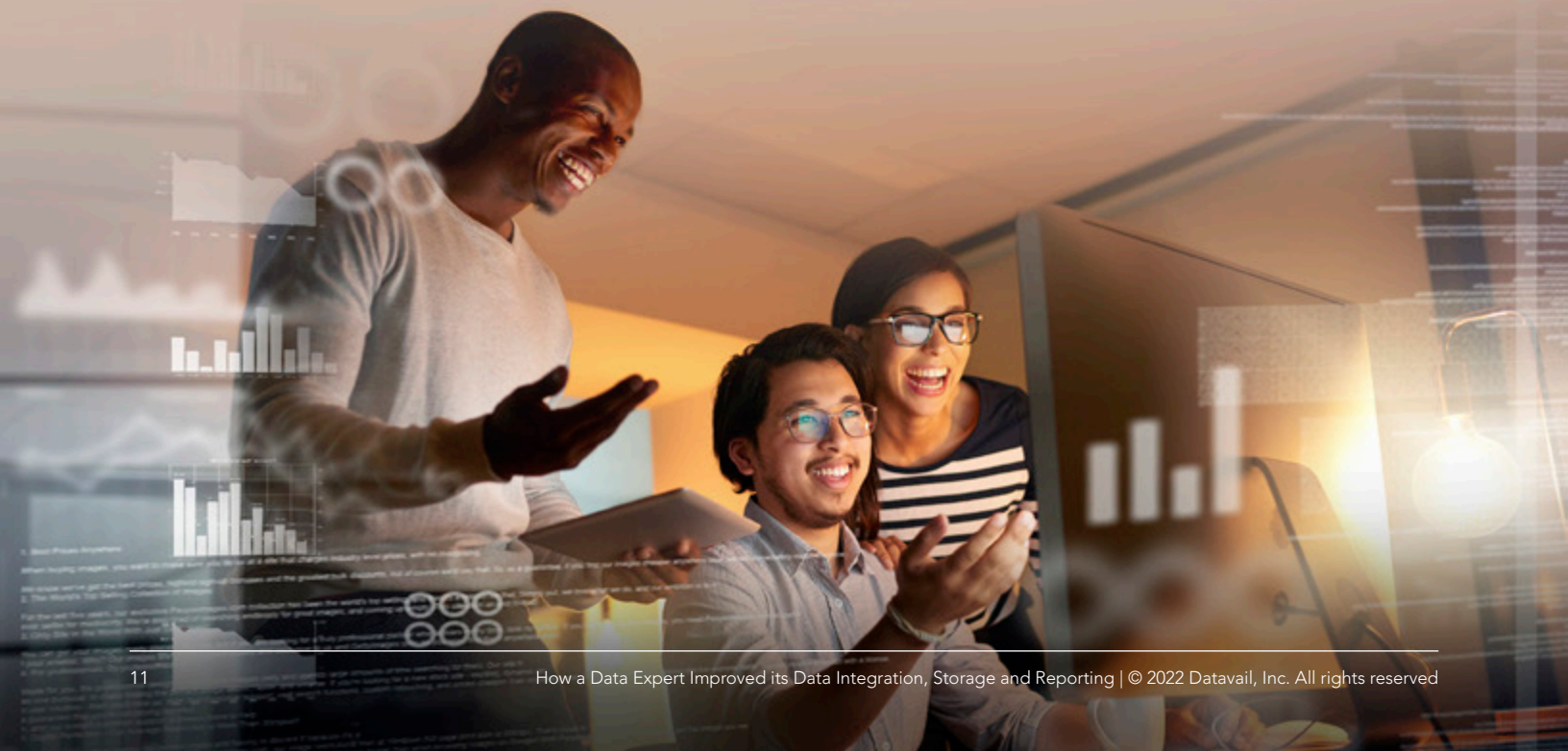
- Datavail's cloud engineers had taken into account and built into the platform all the environmental variables that define the Client's specific processing requirements, including:
    - state-by-state specificity,
    - verification and validation of individual state and conglomerate data,
    - overarching system functionality, and
    - audit, balance, and control (ABC) procedures.
- They had consolidated multiple on-prem tools into a few cloud-based data management engines that provided significantly more service, security, and control than the Client could achieve in its on-prem organization.
- The Power BI suite provided data set generation that facilitated specific, relevant, and granular insights into whole corporate concerns, individual state assets, and the activities of individual customers. What most impressed the Client here was Datavail's elimination of redundancy. Prior to the project, the Client's on-prem architecture generated numerous redundant reports or reports that included excessive redundant material. Having control over report format with the confidence that appropriate data were populating those fields assured the Client that its data, its insights, and, consequently, its decisions would be made based on the best possible corporate intelligence.
- Not least significant, the modernization project optimized the Client's budget, making it both cost-effective and an excellent return on the migration investment.

The Consequence

In the end, Datavail more than proved its concept and the value that it offered to the Client.

- The tech company responded to the most immediate concern: supplying the technological talent and expertise to this Client's particular technology challenges. By providing both administration and development services and supports, Datavail ensured that the Client had the professionals available to maintain their regulatory compliances, as well as the database administration capacity to ensure that their corporate information was accurate and reliable.
- Datavail's comprehensive knowledge of ADF and Power BI also assured the Client that the proposed tools were both appropriate and optimal to achieve their requirements. With these tools in place, the Client now knows that its current customer base is well-served and can safely contemplate growing that base by offering equally valuable services in the future.
- This Client was particularly impressed with the economic value of the Datavail solutions. Despite its global reach, it is a small organization and needs to maximize its budget as much as possible. They were pleased with the highly polished technical skillsets provided by Datavail at a competitive price they could afford.

Datavail's cloud computing consultants and experts are available to serve in many roles, from being a trusted vendor to pairing up as a strategic technology partner. Their mastery of ADF, Power BI, database administration, application development, analytics, and more provide their clients with the technological support and advances needed to thrive in today's accelerated global economy.



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