

A photograph of the Tarleton State University campus. In the foreground, three tall black flagpoles stand on a green lawn. From left to right, they fly a purple flag with the university's seal, the Texas state flag, and the United States flag. To the right, a stone pillar with 'TARLETON' written vertically in metal letters stands next to a brick building. A yellow globe sits atop the pillar. In the background, there are trees and a parking lot with several cars. A flower bed with purple flowers is in the bottom foreground.

EMBRACING THE FUTURE: Tarleton State University's Data Driven Transformation

Embracing The Future: Tarleton State University's Data Driven Transformation



Tarleton State University was facing the challenge of siloed data and a lack of centralized reporting across various data sources. With new leadership focused on implementing a long-term, data-driven growth strategy called Tarleton Forward 2030, the university needed a comprehensive solution to support informed decision-making. This case study outlines how Columbia Advisory Group (CAG) created and implemented a Microsoft Azure-based data warehousing and reporting solution to address these challenges and drive Tarleton State University's data-driven transformation.

The University's Data Challenge

Vision and Focus

Tarleton has a capable IT team for operations. There is an excellent Institutional Analytics team concentrated on meeting regulatory requirements. As is common in organizations early in their data analytics journey, the focus, direction, and ability to make Tarleton Forward 2030 happen were not yet established.

No Centralized Data Infrastructure

The university consisted of a combination of on-premises databases, a growth in cloud-based solutions, and multiple reporting software products. Thus increasing the number of data sources and diversity of data. The outcome is that data across systems was distributed across multiple systems, and standard data definitions made reporting inconsistent, and a challenge at best. Extracting data from sources and then manually consolidating into spreadsheets is not uncommon.

Where two systems captured the same data, determining the source of truth of the data is problematic. Data extraction for aggregation from source systems could vary by each system. The other challenge with premise-based systems in the need to procure hardware, and software to meet the peak demand. This can mean provisioning 80% or more than required for day-to-day operations creating unnecessary cost.

No Centralized Data Infrastructure

The university consisted of a combination of on-premises databases, a growth in cloud-based solutions, and multiple reporting software products. Thus increasing the number of data sources and diversity of data. The outcome is that data across systems was distributed across multiple systems, and standard data definitions made reporting inconsistent, and a challenge at best. Extracting data from sources and then manually consolidating into spreadsheets is not uncommon.

Embracing The Future: Tarleton State University's Data Driven Transformation

Where two systems captured the same data, determining the source of truth of the data is problematic. Data extraction for aggregation from source systems could vary by each system. The other challenge with premise-based systems is the need to procure hardware, and software to meet the peak demand. This can mean provisioning 80% or more than required for day-to-day operations creating unnecessary cost.

Data Governance

The data silos across the university lacked structured governance. This arrangement hindered data aggregation, determination of causal factors, and extracting of valuable, workable insights from the data. The lack of common definitions of terms, common data definitions, inconsistent data types or fields introduced discrepancies in aggregation. Discrepancies in data definitions can impact data veracity. In addition, the data teams could not access modern reporting tools without spending undue capital.

Hindered Data Analysis

The combined issues of disparate data sources without governance created issues for analysis. In turn, when a more detailed analysis is required, finding the proper data is extremely challenging. The absence in standardized, aggregated data put the issue back on the department or individual to figure it out on their own. The impact is a decrease in efficiency having to solve issues at the individual level.

Proposed Solution by CAG

Columbia Advisory Group partnered to propose a workable, cost-effective solutions to Tarleton's assimilating and

analyzing data issues. The solutions involved implementing a modern data infrastructure and included the following measures:

Leveraging the benefits of the cloud

CAG proposed that Tarleton leverage the speed and agility of cloud platforms for storing and analyzing their data. Tarleton benefitted from a future-proof cloud platform that could meet current and future information-related requirements..

CAG worked with Tarleton's IT team with focus and direction on their existing cloud skills to shorten the implementation timeframe and leverage existing talent. The benefits of a cloud-based solution include:

- a)dynamic scalability for performance and storage – no over provisioning,
- b)scale systems up or down as needed for cost avoidance, and better expense management,
- c)a central structure for managing privacy and security.

Leveraging Existing Technology Skills

Tarleton University's IT team had prior experience with Microsoft Azure Cloud, thus saving on the expenses incurred and time for training and setting up a new cloud platform. The IT team was well-versed in the Azure cybersecurity protocols, Microsoft Azure Sentinel and Microsoft 365.

Overcoming Data Security Challenges

The primary benefit of working on a known cloud platform was that the IT team was familiar with Microsoft Azure Sentinel and Office 365 for staff, students, and faculty. Such deep pre-existing knowledge removed

Embracing The Future: Tarleton State University's Data Driven Transformation

hurdles concerning security architecture. The university's database team also had considerable experience working with Microsoft based technologies such as T-SQL, Azure, and PL/SQL.

Leveraging Core Azure Tools

The architecture established by CAG was simple and leveraged native solutions where possible:

- The Azure Data Factory extract-transform-load (ETL/ELT) software is used for ingesting data from on-premises and vendor-hosted cloud systems.
- Azure Storage was used as a Data Lake for semi-structured, unstructured, and transient data, while Azure SQL server was used for storing structured data.
- Azure Purview Data Catalog was formed as the foundation supporting data governance. Purview created an up-to-date map of the data landscape with automated data discovery, sensitive data classification, and end-to-end data lineage.
- Azure Application Programming Interface (API) Management Services provided a secure and standard interface for exposing selected data to external partners and applications. This facilitates a common data access mechanism reducing reliance on multiple solutions to access data across the source systems.

What is Data Warehousing?

A data warehouse is a large, centralized repository of data that is specifically designed to support business decision-



-making activities. It collects data from various sources, such as transactional systems, databases, and other sources, and then stores and organizes that data in a way that makes it easier to access and analyze. The data in a data warehouse is typically structured, organized, and optimized for query and analysis, allowing decision-makers to gain insights and make informed decisions based on the data. It is optimized for querying and reporting rather than transaction processing, which is the primary purpose of operational databases.

The process of putting data in a data warehouse is typically divided into four main stages: extraction, transformation, loading, and verification. Here is an overview of each stage:

- **Extraction:** The first stage is to extract data from various sources such as transactional systems, operational databases, cloud based platforms, and other sources. This data is extracted based on predefined rules and criteria.

Embracing The Future: Tarleton State University's Data Driven Transformation

- **Transformation:** The extracted data is then transformed into a format that can be loaded into the data warehouse. This process involves cleaning, aggregating, and integrating the data to ensure its accuracy and consistency.
- **Loading:** The transformed data is loaded into the data warehouse, which involves moving it from the staging area to the target database. This can be done using different loading techniques, such as full or incremental loads.
- **Verification:** Finally, the loaded data is verified to ensure that it has been loaded correctly and that it meets the quality standards required by the organization. This involves validating data completeness, accuracy, and consistency, and resolving any issues that may arise during the verification process.

Data warehouse metadata is information that describes the structure, contents, and usage of a data warehouse. Metadata is used to manage and control the data warehouse, as well as to support data integration, data quality, and data governance.

- **Data dictionary:** A data dictionary is a document that provides a description of the data elements used in the data warehouse, including their names, definitions, and relationships.
- **Data lineage:** Data lineage is a record of the history of data as it moves through the data warehouse, showing where the data came from, how it was transformed, and where it is stored.
- **Data models:** Data models are diagrams that show the structure of the data warehouse, including the tables, fields, and relationships between them.



- **Business rules:** Business rules are the policies and procedures that govern the use of data in the data warehouse, including data security, privacy, and access.
- **Data profiling:** Data profiling is the process of analyzing the data in the data warehouse to understand its quality, completeness, and consistency.
- **Data integration:** Data integration metadata is used to manage the process of extracting, transforming, and loading data into the data warehouse

Adopting a Multi-Staged Approach

Any change in work methodology requires time for stakeholders to get used to the new functions. To reduce users' challenges, Tarleton University focused on implementing a multi-staged implementation approach to creating a centralized data warehouse. The staged approach enabled its staff to learn operations without becoming overwhelmed. It also empowered the use of the new solution immediately without devoting stakeholders' time to learning new tools.

Embracing The Future: Tarleton State University's Data Driven Transformation



The First Stage: Creating an Azure Infrastructure

The Azure infrastructure was established, and quickly configured ingestion of the required core datasets. The initial data included Ellucian Banner and Degree Works data from the Oracle-based, on-premises servers. In addition, ingestion jobs were created for cloud-based data sources from Salesforce, Target-X, and Canvas LMS. The Azure Data Factory creates data pipelines to securely access each system with scheduled ingestion jobs from a single management tool.

The Second Stage: Reproducing the Reported Datasets

Initially the solution reproduced the reporting data sets on-premises in an overnight process by copying the Ellucian Banner database. However, this procedure takes time and effort. The second phase improved the process by updating data sets in Azure and then reproduced the specific reporting (analytics) data sets used by Institutional Analytics. This process saved time and created access to the

data via the cloud, and common reporting tools.

The Final Stage: Making Data Available for Analytics and Reporting

The end-user can access the data and metadata providing information about the structure and content of the data in the data warehouse. This reduces the reliance on a technical person or DBA to locate data, and relationships. Additionally, the metadata can provide insight and prioritization of additional data sets to add to the ingestion process to aggregates information for analysis.

Summary

Columbia Advisory Group, working with Tarleton State University, helped design information systems around protecting the confidentiality, integrity, and availability of their valuable data assets. University data is now securely available for analysis and making informed data-driven decisions. The university can now let its data inform decision-makers about what is happening and what they can do to help students, retain students, and improve their experiences. *Tarleton Forward* also empowers students to stay students for their educational careers and helps the university understand what happens after departure and how to stay connected to alums.

Tarleton State's new data warehouse solution is realizing the following benefits:

Cost reduction: The cloud-based, centralized data warehouse solution required a fraction of this cost versus an on-premises solution. The university's team could manage the system without additional staffing requirements.

Quick implementation: The initial implementation



took five weeks to complete.

Improved access to data: Tarleton's Institutional Analytics team realized quick access to correlated data. Further, Tarleton is able to store recorded data over time. This can be useful for trend analysis, forecasting, and identifying patterns and anomalies over time.

Enhanced and Meaningful reporting: The reporting capabilities expanded the types of reports available for analysis. There is an additional benefit by providing a concrete platform for future AI/ML work.

Improved Mobile App: The Azure's API management services provide secure access to selected data for partners and internal use. One key benefit is the ability for the Tarleton mobile app to provide users with a better personal and customized experience.

Centralization of data: The centrally-accessed and secured cloud-based data services provide access across all Tarleton campuses.

Leveraging automation capabilities: Manual tasks requiring datasets for programs that took 6-8 weeks for students could now be automated. This reduced human effort, and time commitment.

Deep analysis: The solution enables the deep analysis and correlating of trends and unstructured/semi-structured data to develop prediction capabilities based on the enhanced data sets.



Innovate. Optimize. Secure.

17950 Preston Rd
Suite 380
Dallas, TX. 75252

www.columbiaadvisory.com