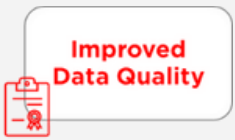
ADASTRA

Success Story: Data Ingestion Readiness Framework and Pilot for a Pharmaceutical Company



Background and Challenge

Our client, a global pharmaceutical company, was planning to move from their existing legacy systems to a more modern architecture on Azure. Prior to the migration, they wanted to get a better understanding of their data sources, the different transformations, and the lineage of those transformations. They were looking for a framework that would not only describe the process of migration but would continually help them discover data and understand data lineage.

The client was looking for a partner that could establish a Data Ingestion Readiness framework for analysis and documentation of all their data sources and transformations that would get ingested into their Azure Data Warehouse. They also wanted the partner to initiate a detailed data mapping and discovery exercise based on this framework.

Adastra was one of the potential partners the client reached out to for the project, and based on our detailed proposal, we were selected as the vendor of choice.

Adastra's Approach

Before the migration would be initiated, an investigation of the current state was needed, and the approaches for the target state needed to be defined. This required an assessment and cataloguing of the data repositories and metadata, and of the transformations used between the various layers of data repositories.

The data repository discovery would drive the potential repository redesign and implementation in Azure. Similarly, the transformation discovery, or reverse engineering results, would drive the transformation logic design or redesign and implementation in Azure. The discovery process was an input to the report design and creation in Azure. Changes implemented in the data repository would impact both the transformation code design and the report design, since it is the target of the transformation code and the source for the reports. Also, the implementation of the data repository had to be completed before the code could be implemented or reports were created.

DISCOVERY

Adastra's experts undertook a current state assessment of the 8-10 data sources that were in scope. Our objective was to understand their current architecture, outline best practices for profiling of their data and preparing it for migration, and recommending a future target architecture encompassing all the data sources. The output of this process was a documented set of data definitions, profiles, and data analysis findings.

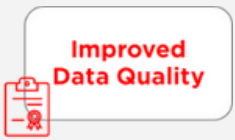
The Discovery process included the following:

ABOUT THE CLIENT

Industry: Pharmaceuticals

Region: Global

Success Story: Data Ingestion Readiness Framework and Pilot for a Pharmaceutical Company



Data Cataloguing of sources, entities, and attributes:
The first step was to define the data scope and locate related data. We also found existing documentation describing the data, to augment and complement the documentation and profiling and present as complete a picture as possible

.

Data Dictionary Documentation:
Using the data scope as a starting point and with inputs from SMEs, we created and documented data definitions.

Data Profiling and Data Quality Analysis:
Profiling was done on all the source and target tables in scope, and the results were captured and documented in a centralized location, along with data definitions collected from the SMEs. The information collected at this stage would later be used to define the required tables and create proper structures in Azure.

A Data Quality analysis was undertaken to develop a better understanding of the source data, identify data quality issues, determine whether the source data satisfies target data rules and constraints, and to provide input for identification of gaps in the source data. This also helped us identify required code translation from source to target. The DQ analysis was done at field and record level to ensure that the data was accurate, complete, consistent, and validated against business rules, and to remove duplication and invalid values. We started by identifying field level and record level data quality issues and reviewed identified issues with the client's SMEs to gather their feedback on validity and business impact of these issues. Finally, we documented the feedback and the derived business rules and data validation rules.

Data Lineage:
A process was also defined for lineage discovery manually and by using accelerator tools to trace the data flow and document the data lineage mapping for each source/target interaction.

The data lineage mapping helped the client understand the origin and life cycle of their data and the data flows. With a clear picture of where the data moved over time, we were able to determine which downstream applications would be affected by changes in the data.

DATA INGESTION READINESS FRAMEWORK AND ROADMAP

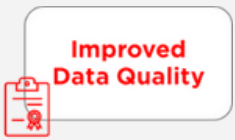
The final Data Ingestion Readiness Framework captured an overview of the client's data sources, patterns for ingestion, approach and templates to catalogue data sources, entities/files, and attributes, approaches for profiling the data and documenting the data quality, and approaches and templates to catalogue data mappings, transformations, and lineage. It also included an approach for reverse engineering of existing data flows, dependencies, and tooling.

The Framework also detailed out the approach for the target state design, including the data repository design, transformation design, and report design, capturing prerequisites, tasks, and considerations for each of these.

ABOUT THE CLIENT

Industry: Pharmaceuticals
Region: Global

Success Story: Data Ingestion Readiness Framework and Pilot for a Pharmaceutical Company



Once the Data Ingestion Readiness Framework and Data Discovery Documentation were ready, our team developed a detailed roadmap and recommendations based on it to help them with their migration effort.

DATA DISCOVERY PILOT

Based on the Framework, Adastra successfully implemented a data discovery pilot. This, along with the roadmap and recommendations of the Framework, helped the client plan and implement their migration to Azure seamlessly.

Impact

The Data Ingestion Readiness Framework laid a foundation for the client to better understand the end-to-end process, their data flows, and prepare for the migration and ingestion of their various data sources into the Azure Data Warehouse platform.

By using this framework, the client would have in place a process for ensuring high data quality, which in turn, would allow them to have more confidence when ingesting data into Azure. This also offered an opportunity to optimize code, based on the findings from the reverse engineering efforts.

The framework captured best practices, approaches and methodologies for profiling and cataloguing data and transformations, and for assessing data quality and data lineage. While it was an essential precursor for ensuring a smooth migration, it can continue to be used by the client to establish internal best practices for data ingestion.

The Data Discovery pilot that was conducted by Adastra also helped solidify the practical applicability of the Framework, and learnings from the pilot were useful for the client during the migration implementation.

ABOUT THE CLIENT

Industry: Pharmaceuticals

Region: Global

ABOUT ADASTRA

For over two decades, Adastra Corporation has transformed businesses into digital leaders, helping global organizations innovate, achieve operational excellence, and create unforgettable customer experiences, all with the power of their data. At the forefront of Artificial Intelligence, Data, Cloud, Digital and Governance services, Adastra delivers solutions to enterprises to leverage data they can control and trust, connecting them to their customers – and their customers to the world.

www.adastragrp.com