

Automated Data Management for a Record Warehouse

Application of Automated Data Management for a record warehouse for public records

Who is the client?

The client is a provider of information sources and a risk solutions service provider. The client runs an online platform that aids businesses and public procure client's services in addition to on-site consultation.

What was the client's need?

The need was to digitize client's complete record warehouse which comprised of both historical and recent records. There were almost 50,000 records which were earmarked as relevant and valuable, thus needing digitization and safer storage.



What services did Data-Core provide?

Right at the outset, it was imperative that the digitization process be automated to ensure consistency, speed, and low error pilferage. The project was planned to be completed in four phases, namely - Scanning, Data Extraction, Data Processing, and Persistence.

Scanning - The record hard copies were collected in batches with unique serial numbers for easier reconciliation after every batch processing. The records were scanned using an industrial scanner and the scanned images were uploaded into a Blob Storage on Azure Cloud. Azure Cloud was chosen for its strict compliance with the security guidelines outlined by the US Federal Agencies.

(Continued on next page)

Automated Data Management for a Record Warehouse

Data Extraction - Data-Core chose to use the Microsoft Power Automate (earlier Microsoft Flow) to fully automate the process from fetching the records from the Blob Storage to extracting data to running data processing and then pushing the processed data to a SQL Server Database. The data extraction flow starts with fetching the records from the Blob Storage and then applying the Azure Form Recognizer Cognitive API to extract the image data. The extracted data were pushed into an Azure Table Storage, which is a schema-less storage for no-SQL data. This extraction of business logic is handled from an Azure function which is triggered as the first task in the automation flow.

Data Processing - Data Processing was arguably the trickiest piece of the puzzle. The extracted data had to be sanitized before it could be persisted in a database. The data in the table storage were put into processing using the Trifacta Data Wrangler. Data-Core and the client's officials worked ceaselessly to identify the mistakes in the mistyped or miswritten forms and engendered recipes to rectify the data. The processed data were sent back to the Azure Table Storage to be transformed and loaded to a relational database in the next task.

Data Persistence - Now that the data were processed, the only piece left was to transform them into a structured form and save in a database. This ETL flow was performed in the Azure Data Factory where a pipeline was devised to fetch the NoSQL data from Table Storage, transform them into relational table structures and persist them into Azure SQL Server Database.

The whole process was performed on the Azure Cloud which ensured data security, data persistence, data availability, and transparency, thanks to the SLAs guaranteed by the off-the-shelf cloud service provider.

In Summary

The whole flow takes about 20 minutes to complete for a typical batch of 1,000 - 1,500 records. The development and the project's first phase were completed in about a month's time. The automation expedited and streamlined the whole process which otherwise would have taken several months to accomplish a task of such scale. The framework saved thousands of dollars for the client by slashing the project cost by at least a half of the originally projected workload.



Discover the Data-Core Advantage.

Learn how we can help your company be more relevant, effective and efficient.



USA Headquarters
1500 JFK Blvd., Suite 624
Philadelphia, PA 19102
Tel: 215 243 1990
www.datacoresystems.com

Bristol, USA
111 Sinclair Road,
Bristol, PA 19007
Toll Free: 877 300 9529
Tel: 267 569 0800

Las Vegas, NV
1771 East Flamingo Road Suite B100
Las Vegas, NV 89119
Toll Free: 877 300 9529
Tel: 702 795 9559