

CASE STUDY

Data Orchestration for a Hearing Aid Company Looking to Transfer Large Files



OVERVIEW

Headquartered in Minnesota, US, the client is a well-established hearing aid company operating more than 1,500 franchised locations nationwide.

CHALLENGES

The client encountered difficulties in manually transferring large files, following data processing operations from distributed compute nodes. Additionally, the integration point generated multiple response files, each requiring download and processing for meaningful interpretation. The existing Azure service component faced limitations, preventing file writing on the SFTP location. Furthermore, the external SFTP lacked trigger or callback functionality for notifying the current service components when response files were ready for transfer. Continuous polling of SFTP was deemed cost-prohibitive and unreliable.

INDUSTRY

- Franchise
- Hearing Aids

Data Warehouse

- Data Orchestration
- File Transfer Services

Azure Data Factory

- Azure Logic Apps
- Azure Machine Learning



SOLUTION

We devised a precise and fault-proof solution by leveraging various components within the Azure ecosystem. Using Azure Data Factory orchestration pipelines, we facilitated the transfer of CSV files from Azure blob storage to the IMDATA FTP location via Azure Logic App. After processing the files, the output files were transferred back to Azure blob storage. The solution was implemented in three distinct phases:

- Implementation of a dynamic parametrized Logic App for transferring source files to the integration point SFTP. This Logic App is called by the current orchestration pipeline as the next step following the completion of distributed node computation.
- Deployment of a separate orchestration pipeline within Azure Data Factory. This pipeline utilizes a Logic App to check for response files sent by the integration server at the SFTP output location. The smart retry polling policy determines intervals based on a machine learning model trained on previous file sizes and response times. The system builds an exponential retry process until it detects files or raises an alert for non-receipt of response files within a reasonable time limit.
- Execution of file copying from the blob location using Azure Data Factory service.

BENEFITS

- The solution component is fully dynamic and parameterized, applicable in any pipeline for transferring data from any blob location to the FTP location.
- The ability to send and receive files with intelligent polling, eliminating the need for callbacks or hooks.
- Minimization of human efforts, resulting in overall cost savings for operations.
- Elimination of continuous polling requirements, leading to cost savings associated with polling activities.



About Contata

Contata Solutions is a trusted leader in technology and digital innovation. Through our work in data engineering, data analytics, machine learning, marketing automation and app development, we deliver solutions that address complex problems in ways that are simple, insightful and impactful.

Our promise and value proposition to our customers is simple: we leverage our deep technical expertise and global presence to bring software products and data-driven decision capabilities to life.

Founded in 2000, Contata is a privately-held company headquartered in Minneapolis that serves clients globally from offices in the United States and India.