exposé

Exposé Case Study

A configurable ingestion framework supporting files from multiple sources in multiple formats, enabling consistent and scalable loading into the data warehouse.

Education - Generic Data Ingestion Pipeline

CASE STUDY

Education - Generic Data Ingestion Pipeline

A configurable ingestion framework supporting files from multiple sources in multiple formats, enabling consistent and scalable loading into the data warehouse.

PROBLEM

A leading education provider supporting schools across the state needed to ingest data from a growing number of internal and external sources into its enterprise data warehouse. These sources included file-based and system-based integrations such as SharePoint repositories, SFTP locations, and other managed data feeds. The incoming data varied widely in structure and format, including Excel, CSV, and JSON files.

Historically, each new data source required custom ingestion logic tailored to its location and format. This resulted in increased development effort, longer onboarding times, inconsistent ingestion approaches, and delays in making data available for reporting and analytics. As demand for data grew, this approach limited scalability and slowed access to insights.

SOLUTION

A generic, reusable data ingestion pipeline was designed and implemented using Azure Data Factory (ADF) to support data ingestion from multiple source systems and file-based repositories, including SharePoint and SFTP locations.

The solution leverages control tables to drive ingestion behaviour through configuration rather than code. These control tables define source locations, file formats, schema details, and load rules, enabling the onboarding of new data sources with minimal development effort.

Dynamic column mapping handles variations in source data structures, allowing the pipeline to ingest datasets with different schemas while maintaining consistent loading into the data warehouse. This approach supports common formats such as Excel, CSV, and JSON, and ensures flexibility as source structures evolve over time.

By combining metadata-driven control tables with dynamic mapping, the ingestion framework delivers a scalable, consistent, and highly configurable solution that accelerates data onboarding while maintaining governance and reliability.

BUSINESS BENEFITS

- **Accelerated data onboarding:** New sources and file formats can be onboarded quickly through configuration rather than custom development.
- **Reduced development effort:** Reusable ingestion patterns significantly reduce future build and maintenance effort.
- **Faster access to insights:** Data is ingested and made available to the business more quickly, enabling timely decision-making.
- **Improved scalability:** The framework supports scaling to larger data volumes, new source systems, and additional formats without increasing complexity.
- **Improved data consistency and governance:** Standardised ingestion processes improve data quality, traceability, and reliability across the platform.
- **Lower long-term costs:** Reduced rework and technical debt deliver ongoing cost efficiencies.