

Accelerating data ingestion to the modern data platform.

Organisations are continuously adapting to changing customer behaviours and dynamic market conditions. It has become crucial for organisations to be efficient and deliver value to customers. Embracing the latest technologies and undertaking digital transformation journeys has become essential for delivering this value. One crucial factor in achieving this is using data and analytics as key accelerants in digitisation and transformation journeys.

Many organisations face challenges such as poor data availability, siloed data, and lack of transparency in the data process lifecycle. These issues can be resolved by implementing a modern data platform that establishes trust through robust data management and delivers accurate and timely data availability, reduced data silos, and improved governance.

What is a modern data platform?

A modern data platform or a modern data stack, as often referred to, is a suite of cloud-first and cloud-native software applications that enable collection, transformation, cleansing, and analysis to drive evidence-based decision-making and deliver value from data. The quality of the decisions depends on how well data is ingested and integrated.

What is data ingestion?

The process of placing data into a central repository, a storage system for later use, is called Data Ingestion. This central storage system, also known as the raw or staging layer in multiple data platforms, should ensure availability and elasticity. Once the data is moved to the central storage system, it can be moved to other layers and applications for record-keeping and insight generation. However, the completeness and accuracy of the data are not just important; they are critical. Missing, incomplete, and inaccurate data can cause very expensive issues downstream, impacting the trust in the data systems and the success of the digital transformation journey.

How is data typically ingested into the data platform?

Organisations have multiple sources that require ingestion to the cloud-based storage systems. Each source, in turn, consists of several tables, views and other database objects. Data engineering teams typically create an ingestion framework and a set of configuration files, and they employ a data ingestion application to move data into the ingestion layer. Data engineers spend days updating the configuration files and ingesting data table by table. The typical time spent processing a table is three business days, which can increase for a more complex table with an increased number of records.

What is Glide?

Ingrity has developed a framework called Glide, designed to expedite data ingestion to popular cloud data platforms. Glide automates the creation of configuration files by crawling through the source database and utilising its metadata. This allows ingesting the entire database at once rather than spending days ingesting data table by table. The generated configuration files can be deployed to Azure Data Factory, a widely used data ingestion tool in the Microsoft Azure Cloud platform. These files contain details of the tables and columns to be ingested and can be loaded into a database and edited using SQL. Users can update data types, rename columns, specify the mode of ingestion, and choose between full load or incremental ingestion.

The framework has been deployed and road-tested in various customer environments of different sizes, receiving outstanding appreciation from customers.

What are the tangible benefits of Glide?

The estimated effort required to process and ingest each dataset from a source to the raw layer is approximately 3 man-days. For a source with 100 datasets, the total time required for ingestion would be 300 man-days, resulting in a significant cost and delaying the realisation of benefits.

With Glide, processing each data set is reduced from three days to three hours, resulting in approx—70% cost efficiency.

TL; DR

- Data ingestion to a centralised storage system is an initial and critical step in deriving value from organisational data.
- Ingrity's Glide is a robust framework designed to accelerate data ingestion to the modern data platform by automating much of the data ingestion process.
- This streamlined process results in significant cost savings and provides much-needed efficiency and governance.
- The framework is equipped with automated metadata management, and it is fully customisable and extensible to suit each customer's unique use cases.