

Penguin: AI-driven Semi-automatic, No-code Schema Matcher and Mapper

Fresh Gravity's comprehensive AI-driven Semi-automatic Data Matcher & Mapper tool helps enterprises in defining and creating a uniform and consistent global schema from heterogeneous data sources. It combines a novel visual interface with an algorithm-based recommendation engine that helps enterprises accelerate their journey from being data rich to being data driven.

Background

Merging and coalescing data from multiple and diverse sources into different data formats continues to be a critical problem in modern information systems. Schema Matching, the process of matching elements of a source schema with elements of a target schema, and Schema Mapping, the process of creating a query that maps between two disparate schemas, are at the heart of data integration systems.

There are several challenges that enterprises currently face:

- Enterprises today get their data from business partners and other third parties, as well as internal departments. Each of these information systems use their own, specific, and proprietary data model. Additionally, they can arrive in a myriad of formats for e.g., relational, multi-dimensional, object-oriented etc which are also constantly changing. It is crucial to map elements in a syntactically and semantically correct manner and in an on-going basis. It is a constant uphill task to keep pace and adapt your data platform to ingest and get rich insights from these data sources.
- Inconsistent data formats, protocols, query languages coupled with noisy, duplicate and error prone data create a tremendous challenge in creating a unified global schema for today's distributed sources.
- Existing tools and processes struggle to provide seamless data ingestion from third-party data sources. Enterprises have little to no control on the third-party data and its associated metadata. This makes their data engineering pipelines vulnerable and often necessitate a manual, time-consuming intervention.
- Enterprises are stuck with proprietary and inherently rigid data models that were built over years. This results in tightly coupled integration ecosystem that is not interoperable and fails to keep pace. Consequently, enterprises end up with data silos and a very high operational maintenance cost.

About Penguin

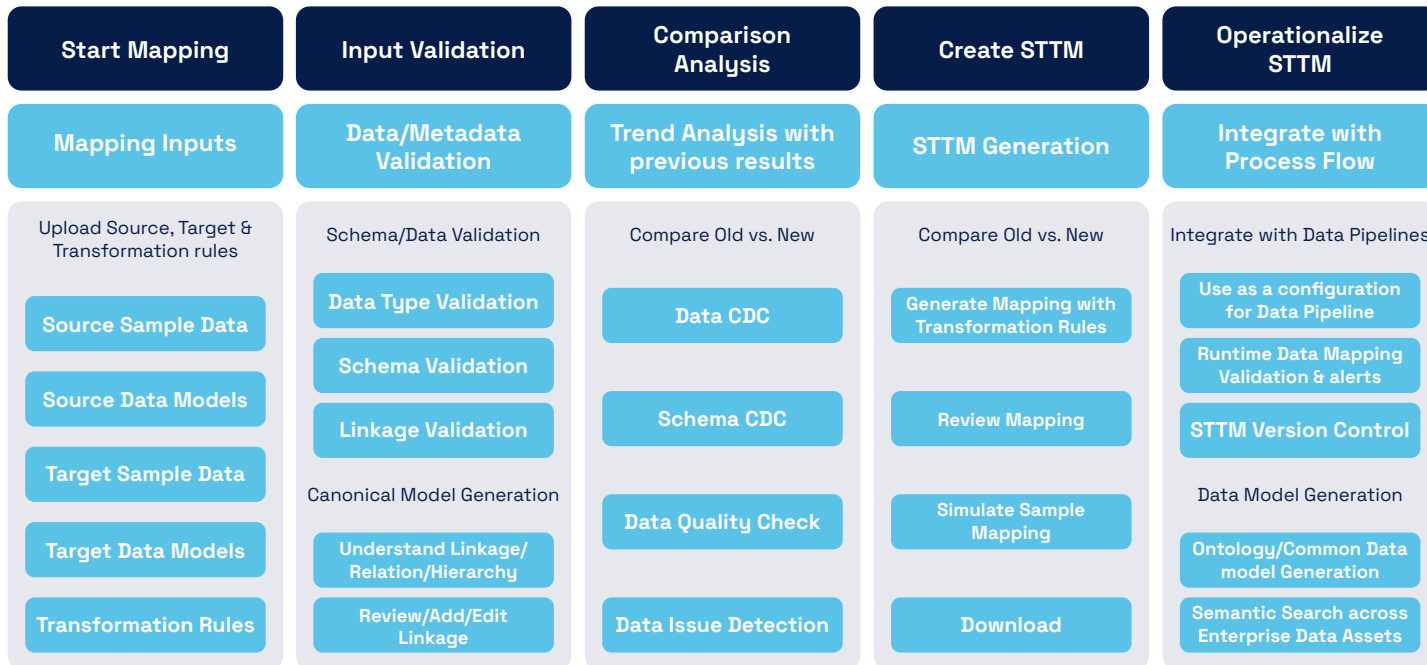
Developed by Fresh Gravity, Penguin is an innovative, best-in-class solution that is driven by knowledge graphs and machine learning. It leverages metadata as well as data in conjunction with our own proven and tested data classification algorithms. Penguin learns semantic information from various data models and data sources. Once it learns and understands the semantic associated with data sources, it automatically generates corresponding ontology. Additionally, it also does automatic contextual mapping using the semantic information.

Challenge	How Penguin Helps	Value Delivered
Data Silos, Rigid and/or no common data model	Penguin's auto mapping & auto translation features enable interoperability between siloed data sources. It creates an enterprise-wide common data model by combining different data sources	Eliminates data silos, democratizes data, and consequently enables enterprises to be Data driven
Time-consuming, manual and error prone ETL process	Penguin provides automatically generated instant source to target mappings with templated ETL scripts that can be integrated with any leading Data integration tool	Creates a nimble framework to access, analyse & quickly deliver insights from raw data. Reduces end-to-end ETL development time by 50-60%
Data ambiguity and lack of context around data assets	Proprietary algorithms based on Natural language processing that enable semantic search across enterprise data assets. Penguin not only considers the data and the associated metadata but also the context for the data making it more intelligent in its recommendations	Enables enterprises to unleash the power of their data and analytics platform by making it more usable
Rigid data integration with third-party data vendors. Change in content or structure of third party provided data requires lot of manual and re-development work	Penguin provides a self-configuring schema matching algorithm that can automatically adapt to the dynamic nature of third-party provided data	Enhances flexibility and provides business continuity while ingesting dynamic third-party data. Eliminates slow and error prone manual work involved in analysing the schema/ data changes

The main aim of the tool is to -

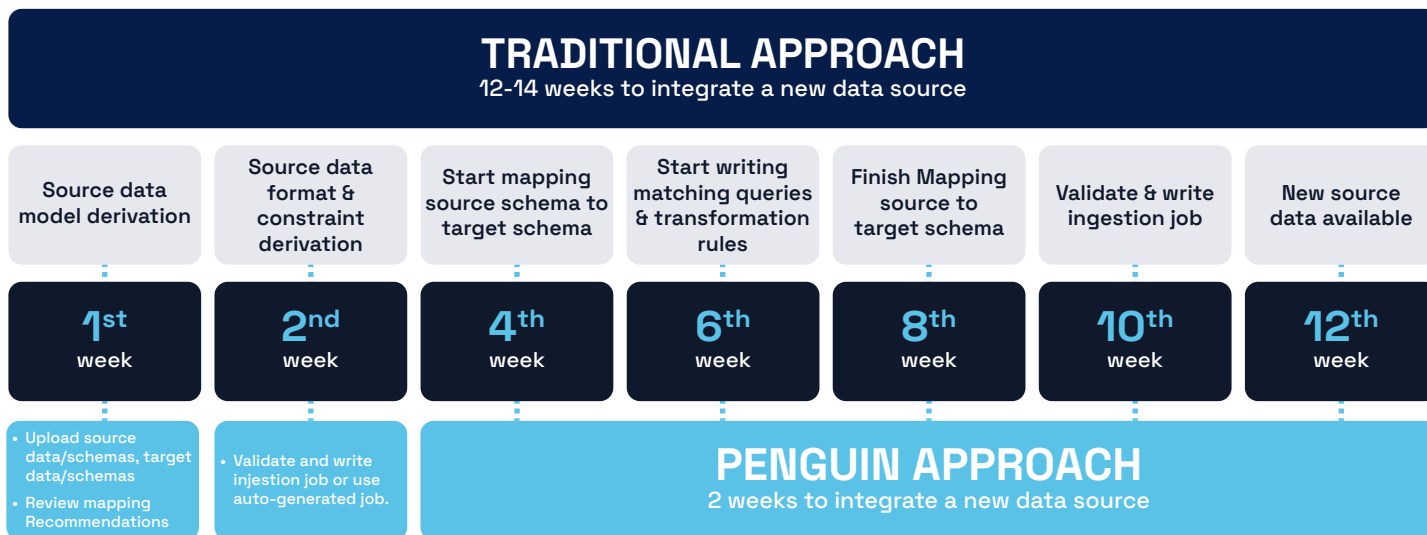
- Eliminate Data Silos and enable enterprise-wide common data model that enables interoperability
- Drastically reduce time to ingest data from a new or updated data source
- Provide a flexible interface with third-party data vendors that reduces dependency on the structure and content of third-party data
- Enable NLP-driven semantic search across enterprise-wide data assets

End-to-End Process Flow with Penguin



Illustrative Example

Given below is an illustrative example of comparative timelines for ingesting a hypothetical new data source with 4-6 tables and approximately 100 odd columns for each of those tables -



About Fresh Gravity

Founded in 2015, Fresh Gravity is an innovation-driven Data & Analytics consulting firm that helps businesses make data-driven decisions. We believe in the transformative power of data using technologies to help clients. To know more about us and our offerings, contact us at info@freshgravity.com.

U.S. Corporate HQ – San Francisco Bay Area

5201 Great America Pkwy
Suite 320
Santa Clara CA 95054

Other Offices:



U.S.A.
Dallas



India
Pune, Bangalore



Canada
Ontario

www.freshgravity.com