

William Mesquita, Markel International, Lenotis Data

12 June 2023

Jonathan Sales, Markel International, UFC

Anna Riepin, Markel International

Fully Automated ETL Using Azure

Markel study case on how to empower Data Scientists



MARKEL



The Cornerstones



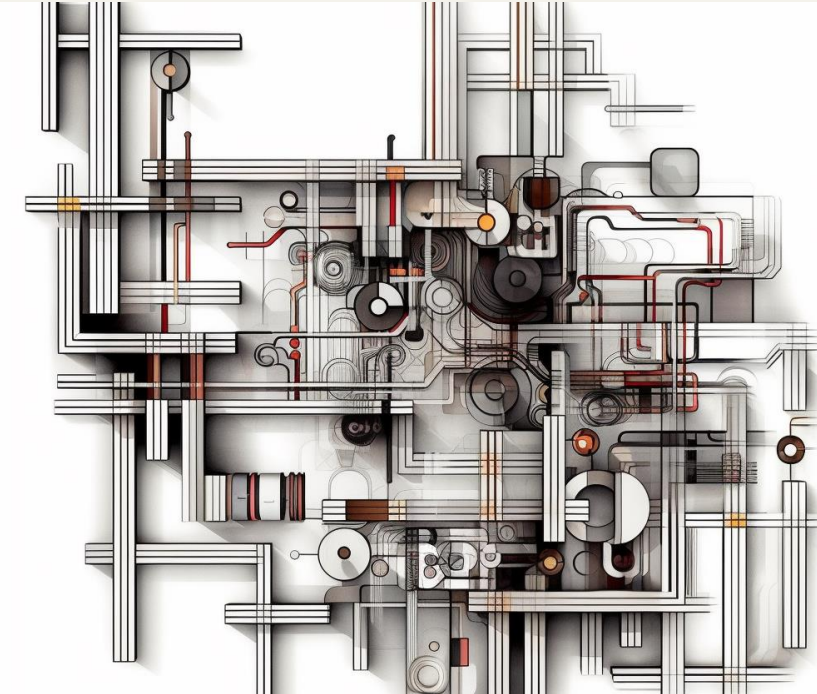
Flexible ETL Framework

A flexible framework allows creating and customising tables quickly, enabling data scientists to manipulate and utilise data from multiple sources according to their unique requirements.



Reliable Data

Garbage in, garbage out. The quality of models directly depends on the quality of data. An effective ETL process ensures data is cleaned and validated, with all the "trash" in quarantine.

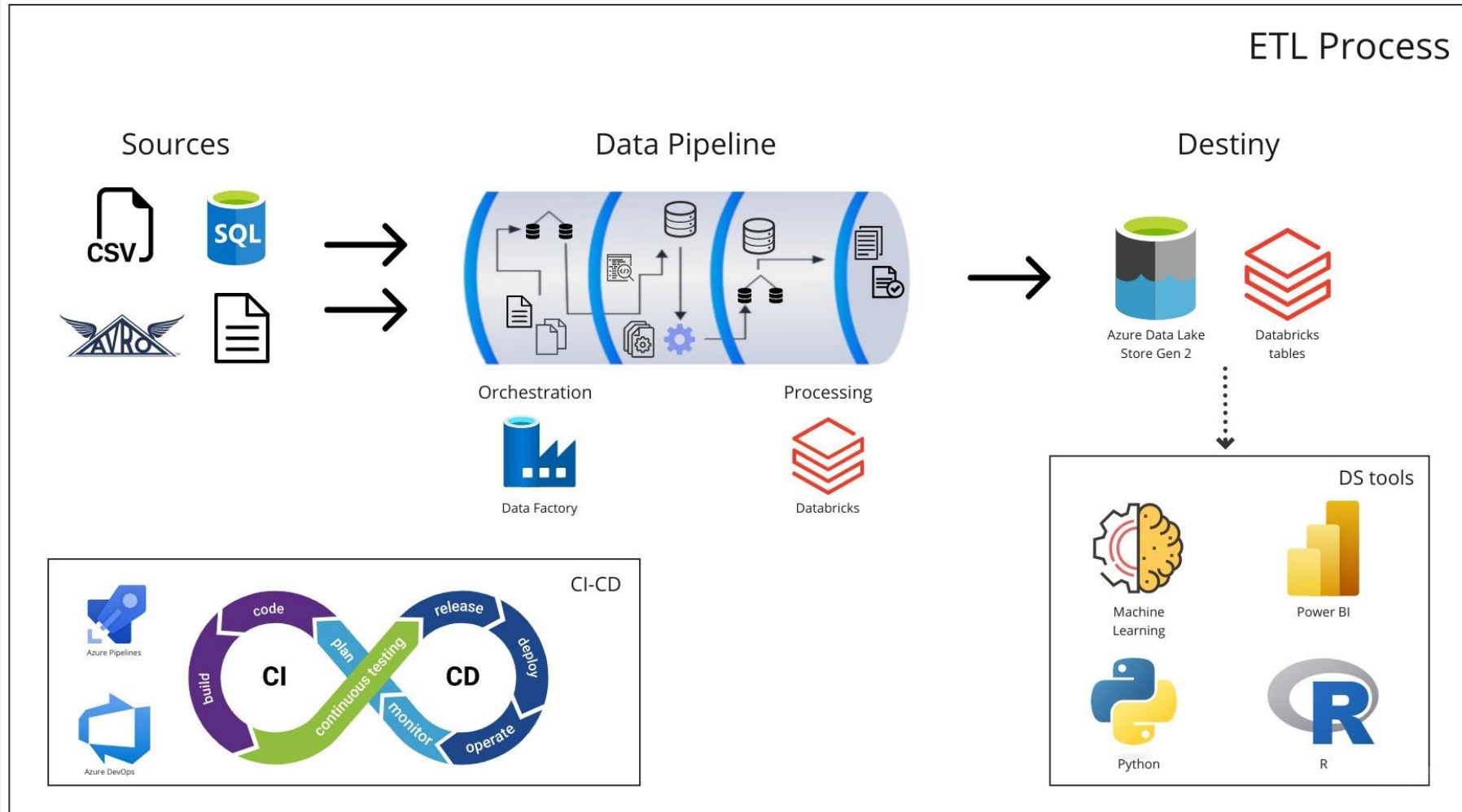


Robust ETL Framework

A robust ETL framework is vital for maintaining uninterrupted, consistent data operations. Supported by automated tests, it mitigates disruptions, prevents unnoticed data quarantine, and safeguards against inaccuracies from transformation changes, thereby ensuring data integrity and reliability.

Market Solution Overview

MINES - MINT Etl Solution





Scalable

Any amount of data
Any refresh window



Robustness

Handle wrong records
Handle temporary issues
Modular approach
Automated Testing



Integration

Multiple Sources
Multiple Formats
Effortless changes



Data Cleaning/Validation

Conform types
Automatic handle locales
Automatic mask PII
Quarantine rejected records



Custom Logic

Custom Clean/validation rules
Custom Transformation logic
Handle Python and R



Collaboration

Users can easily share:
notebooks, code, data, models

Clean, curated data readily
available across company

Voices from Markel: User Testimonials

"The data pipelining, ETL and data lake project has provided the business with a robust, reliable source of data, drastically reduced the number of queries about data accuracy and has hugely increased confidence in our MI. The quality of our discussions has significantly increased, with vastly more time now spent on generating insights and analysis rather than questioning the reliability and quality of our data".

Kevin O'Donovan | Director of NM Data & Analytics

"Canada's participation in Data Lake Project with the implementation of ETL and Quarantine layers have enabled our team to quickly advance and develop Dashboards and Reports to our Business Teams. This toolkit is helping our local Data Analytics team be more efficient in delivering a greater business value to MI Products, as well as , increasing confidence in the data accuracy of our MI Dashboards."

Adriao, Carlos | Data Analytics Product Owner Operations - Canada

"Data lakes provide a centralized and scalable repository for storing vast amounts of raw, conformed, unstructured, and semi-structured data. This consistent format enables me to work with reliable, and integrated data for analysis and modelling purposes".

Halliwell, Chris | Sr. Data Scientist

Connect with Us

- **William Mesquita:** Lead Big Data Developer
- **Email:** DLNMAalytics@markel.com
- **LinkedIn:** www.linkedin.com/in/williammesquita/

- **Anna Riepin:** Senior Data Analyst
- **LinkedIn:** <https://de.linkedin.com/in/anna-riepin-a07410176>

MARKEL