Automated Detection of Sectoral Credit Growth for Banking Risk Management

Fernanda Aristo Abimanyu

Fernanda Aristo Abimanyu

Education

- *Universitas Surabaya (2017–2021)*

Working

- *Data Engineer - NTT DATA (2022–2024)*

Overview Project

- **Web Scraping**
  Scraping of Excel Data from BI's SEKI Website
- **ETL Pipeline**
  Transform Excel into normalized format, compute MoM growth, and detect anomalies
- **Data Warehouse Using Postgres**
  store sectoral credit statistics and risk indicators
- **Interactive Metabase Dashboard**
  showing sectoral trends, credit spikes, and concentration risk

GitHub

- **https://github.com/nandaaristo/Dibimbing-Final-Project**

Project Background

This project implements an end-to-end data engineering pipeline to detect and monitor rapid credit growth across economic sectors in Indonesia, using publicly available data from Bank Indonesia (BI). It supports banking risk management by identifying abnormal credit surges and risk concentration patterns through automated processing and visualization.

The solution consists of ETL processes orchestrated with Apache Airflow, data storage in PostgreSQL, and interactive dashboards with Metabase. It includes anomaly detection logic and sectoral risk flagging to assist financial analysts and regulators in proactive decision-making.

Problem Statement

Banks and regulators require timely and reliable insights into sectoral credit distribution to monitor risk concentration and identify abnormal credit growth patterns. However, sector-level credit data published by Bank Indonesia (BI) is:

- Available only in manual Excel format
- Spread across multiple sheets
- Not structured for analysis or dashboarding
- Updated regularly without an automated integration mechanism

This results in delayed analysis, lack of proactive risk detection, and high dependency on manual processes.

The goal of this project is to build an automated data pipeline that:

1. Scrapes and consolidates sectoral credit data from Bank Indonesia's SEKI website
2. Transforms raw Excel sheets into a unified, analyzable format
3. Calculates Month-over-Month (MoM) credit growth per economic sector
4. Flags abnormal growths as early indicators of credit concentration risk
5. Loads the output into a PostgreSQL data warehouse
6. Visualizes insights through a dynamic Metabase dashboard

The success of the project will be measured through:

Metric	Target
Data freshness	Updated monthly from BI
Risk flag accuracy	Correctly identify ≥90% of large MoM surges
Reduction in manual data processing	>80% automated from source to dashboard
Dashboard	Used regularly by analysts or decision-makers

ibimbing

Data Platform Understanding

Data Source Identification

The primary data source is the official website of **Bank Indonesia**, specifically from the SEKI portal (Economic and Financial Statistics of Indonesia). The dataset used is:

- **File Name**: `TABEL1_5.xls`
- **Content**: Posisi Pinjaman/Kredit Rupiah yang Diberikan oleh Bank Umum dan BPR, berdasarkan sektor ekonomi dan jenis bank.
- **Format**: Microsoft Excel (.xls)

The data represents how credit is distributed across sectors, helping identify sectoral credit growth and liquidity risks

Orchestration

- **Tool**: Apache Airflow
- **Definition**: A workflow orchestration tool to automate and schedule ETL processes
- **Usage**: Handles scraping, transformation, and loading processes using **dynamic task mapping and multiprocessing**

Storage

- **MinIO**:
  - **Definition**: Local object storage system compatible with Amazon S3
  - **Purpose**: Stores raw Excel files and processed `.parquet` files by date
- **PostgreSQL**:
  - **Definition**: A relational database system (RDBMS)
  - **Purpose**: Stores the final transformed and analyzed data used for visualization

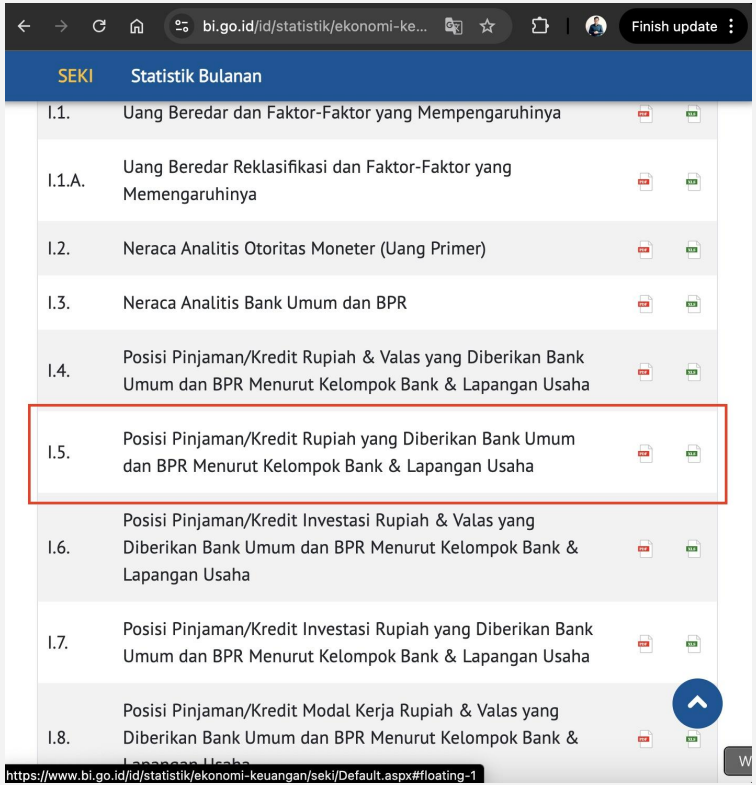
Transformation

- **Tools**: Python, pandas
- **Function**:
  - Merges 3 Excel sheets into a unified format
  - Cleans and reshapes the data
  - Calculates Month-over-Month (MoM) credit growth
  - Flags abnormal spikes for risk monitoring
- **Output**: Cleaned `.parquet` files and PostgreSQL tables (`fact_credit_growth`)

Visualization

- **Tool**: Metabase
- **Function**: Connects to PostgreSQL to visualize:
  - Sectoral credit trends
  - Anomaly detection (MoM > 30%)
  - Credit distribution and concentration across economic sectors

Data Understanding



Data Source and Description

The data originates from the official **Bank Indonesia (BI)** SEKI portal (Statistik Ekonomi dan Keuangan Indonesia). The specific file used in this project is:

- **File Name**: `TABEL1_5.xls`
- **Title**: Posisi Pinjaman/Kredit Rupiah yang Diberikan oleh Bank Umum dan BPR, berdasarkan sektor ekonomi dan jenis bank.
- **Data:** ~ 150 Record
- **Format**: Microsoft Excel (.xls)
- **Access Method**: Public URL (scraped programmatically via `requests`)

Data Structure & Sheets Used

The Excel file contains multiple sheets. This project uses three main sheets:

Sheet Name	Description	Use in Project
`I.5_1`	Bank Persero and Bank Pemerintah Daerah	Used for aggregation and sector analysis
`I.5_2`	Bank Swasta Nasional and Kantor Cabang di Luar Negeri	Used for aggregation and sector analysis
`I.5_3`	Bank Pengkreditan Rakyat	Used for aggregation and sector analysis

Each sheet contains:

- Rows representing economic sectors (e.g., Agriculture, Manufacturing, Services)
- Columns representing monthly credit values, labeled by date
- Credit values in billions of Indonesian Rupiah (IDR)

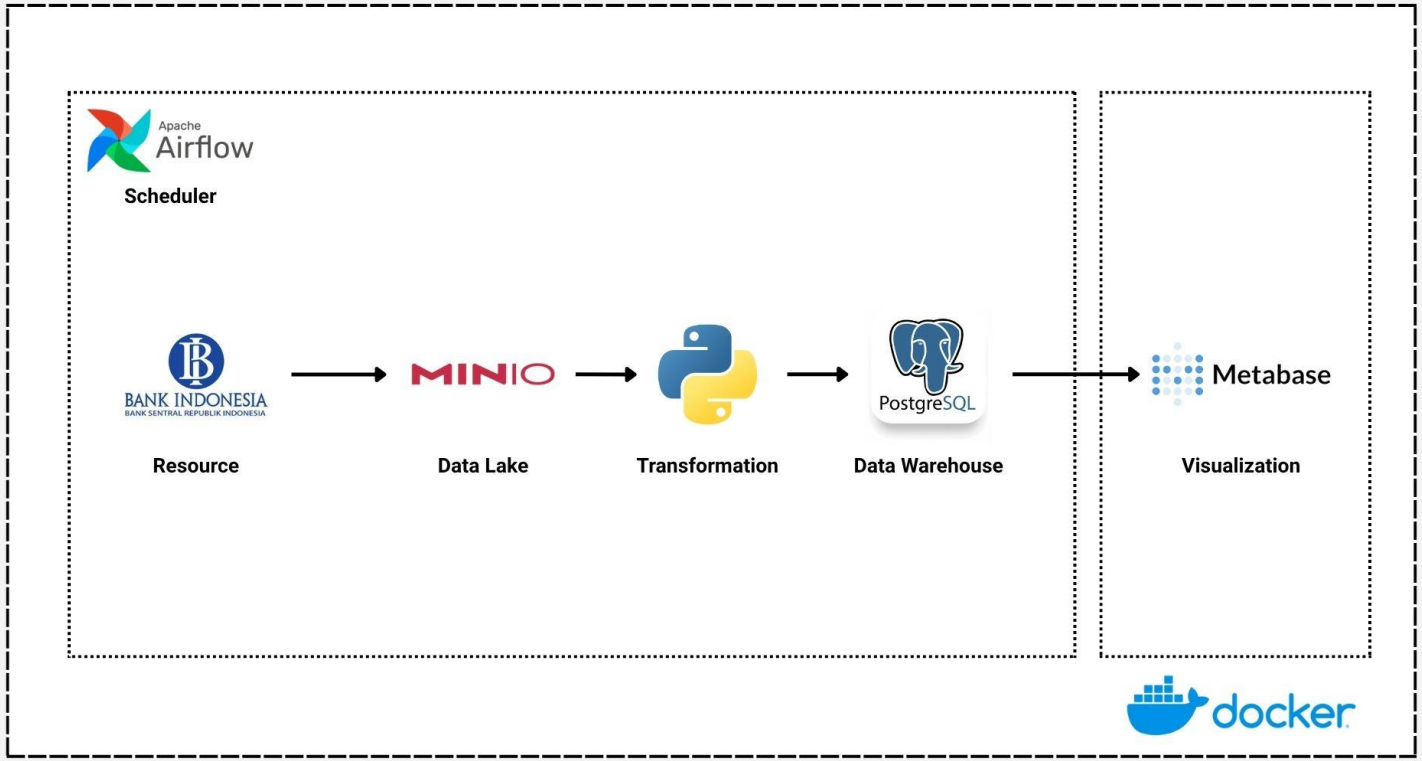
Data Collection Method

- The Excel file is scraped using Python's `requests` library and stored directly into a **MinIO object storage** as the data lake.
- Filename is versioned by scrape date (e.g., `TABEL1_5_20250704.xls`)
- Raw files are preserved in the folder: `kredit-data/raw/YYYY-MM-DD/`

Data Quality Review

- Provided by a **trusted and official** source (Bank Indonesia)
- Structured format (tabular, consistent across periods)
- Timely updates and historical backfill

Transformation & Consideration


Apache Airflow
Scheduler
BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA
Resource
MINIO
Data Lake
Transformation
PostgreSQL
Data Warehouse
Metabase
Visualization
docker

Data Modeling (Business)

Tabel Fakta (`fact_credit_growth`)

- Berisi data kredit bulanan per sektor
    - `kategori` (Kelompok Bank)
    - `sektor` (Sektor Kredit)
    - `bulan`
    - `nilai_kredit`
    - `mom_growth` (analitik)
    - `flag_risiko` (hasil deteksi anomali)

Terima Kasih.