

Proyecto Integrado V - Línea de Énfasis (Entrega 1)

Sergio Andres Rios Gómez
Edwin Alexander Ibarra Ortiz

Recolección Automatizada de Datos Históricos: Taiwan Semiconductor Manufacturing Company (TSM)

Contenido

Recolección Automatizada de Datos Históricos: Taiwan Semiconductor
Manufacturing Company (TSM) 2

 Resumen 3

 Introducción..... 4

 Metodología 5

Resumen

Este informe presenta el desarrollo e implementación de un sistema automatizado para la recolección continua de datos históricos del comportamiento bursátil de Taiwan Semiconductor Manufacturing Company (TSMC). Utilizando técnicas de programación orientada a objetos (OOP), se diseñó un módulo que permite extraer, almacenar y mantener actualizados los datos sin sobrescribir el histórico. El sistema está desplegado mediante GitHub Actions, permitiendo la ejecución diaria de la recolección a las 00:00 horas. La información persiste tanto en un archivo CSV como en una base de datos SQLite, facilitando su trazabilidad y análisis posterior. Esta solución integra prácticas de ingeniería de software como control de versiones, logging de eventos, validación de duplicados y modularidad del código. La recolección de datos económicos relevantes como el comportamiento financiero de TSMC permite sentar las bases para análisis predictivos, estudios de mercado y evaluación de riesgos.

Introducción

En la actualidad, la automatización de procesos relacionados con la recolección y análisis de datos financieros es una competencia esencial en el ámbito de la analítica económica y el desarrollo de soluciones inteligentes de negocio. Particularmente, el monitoreo continuo de indicadores económicos permite generar modelos de predicción, evaluar escenarios y apoyar la toma de decisiones informadas en entornos dinámicos.

Taiwan Semiconductor Manufacturing Company (TSMC) es considerada una de las compañías más influyentes en la industria global de semiconductores. Su comportamiento financiero es un indicador clave para evaluar la estabilidad y evolución tecnológica del sector. Por tanto, disponer de un histórico confiable y actualizado de su actividad bursátil resulta fundamental para investigadores, inversionistas y analistas financieros.

Este proyecto tiene como finalidad automatizar la recolección de datos históricos de TSMC desde Yahoo Finance, asegurando su persistencia en formatos estructurados y la trazabilidad de cada ejecución del sistema. Se ha adoptado un enfoque técnico basado en principios de programación orientada a objetos y prácticas de ingeniería modernas, como la integración continua mediante GitHub Actions, para garantizar la calidad y mantenibilidad del sistema.

Metodología

El desarrollo del sistema se llevó a cabo siguiendo un enfoque modular y orientado a objetos, dividiendo la solución en componentes específicos con responsabilidades bien definidas. Las etapas principales de la metodología fueron:

1. Selección del indicador económico

Se eligió el comportamiento bursátil de Taiwan Semiconductor Manufacturing Company (TSMC) por su relevancia económica e industrial. La fuente seleccionada para la extracción de datos fue Yahoo Finance, utilizando su endpoint oficial para obtener la serie histórica diaria.

2. Diseño del sistema

Se estructuró el sistema en varios módulos:

- collector.py: clase principal encargada del proceso de recolección, validación, almacenamiento y exportación de datos.
- logger.py: configuración centralizada del sistema de logs para registrar eventos, errores y resultados.
- infrastructure/manageDataBase.py: gestión de la base de datos SQLite, creación de tablas y control de duplicados.
- service/queryYahooAdapter.py: adaptación del endpoint de Yahoo Finance y envío de solicitudes.
- service/parseDataYahooQuery.py: procesamiento y estructuración de la respuesta recibida en formato JSON.

3. Persistencia de datos

Los datos recolectados se almacenan de forma redundante en dos medios:

- Base de datos SQLite (historical.db), para facilitar consultas analíticas.
- Archivo CSV (historical.csv), útil para visualización rápida de todos los datos almacenados.
- Archivos CSV de consulta diaria con su respectiva fecha

Se implementó una validación interna que impide duplicar registros si ya existen en la base de datos, garantizando un histórico limpio y confiable.

4. Automatización con GitHub Actions

Se configuró un archivo de flujo de trabajo (.github/workflows/update_data.yml) para automatizar la ejecución del recolector todos los días a las 00:00 horas.

Adicionalmente, la acción se ejecuta automáticamente ante cada push o pull request a la rama principal, facilitando el monitoreo continuo y actualizaciones rápidas.

5. Control de versiones y documentación

Todo el proyecto está alojado en GitHub, con control de versiones habilitado para garantizar la trazabilidad del código y facilitar su evolución. Se incluye además un archivo README.md detallado con instrucciones de instalación, uso y contribución.