A Business Intelligence (BI) Tool for Data Visualization on top of Designed and Implemented DW for a selected open-source BI tool to visualize information for algorithmic trading systems.

Annie Martina Viju
300210450
Data-Analysis
University of the Fraser Valley
annie.viju@student.ufv.ca

Anupam Sharma
300208103
Data-Analysis
University of the Fraser Valley
anupam.sharma@student.ufv.ca

Kripal Kaur
300196082
Computer Information System
University of the Fraser Valley
kripal.kaur@student.ufv.ca

Shanmukha Sree Veda Tippavajhala
300210776
Data-Analysis
University of the Fraser Valley
shanmukha.tippavajhala@student.ufv.ca

Simranpreet Kaur
300200432
Computer Information System
University of the Fraser Valley
Simranpreet.kaur5@student.ufv.ca

Abstract

In the realm of financial markets, Business Intelligence tools play a crucial role in processing vast amounts of unstructured data from diverse sources, aiding in informed decision-making. The primary objective of this project is to seamlessly integrate a chosen open-source BI tool with a Data Warehouse, enabling stakeholders to visualize essential data related to algorithmic trading systems. Through BI visualization, stakeholders can gain deeper insights into market trends, operational efficiency, and revenue potential. The integration of DW with BI facilitates the generation of actionable insights, empowering strategic decision-making and the identification of emerging opportunities within the financial market. This paper discusses the development and integration of a data warehouse (DW) with a Business Intelligence (BI) tool to visualize critical information for algorithmic trading systems.

Keywords *– Data Warehouse (DW), Business Intelligence (BI), Extract Transform Load(ETL), Visualization,*

I. Introduction

Business Intelligence (BI) tools have become indispensable for modern enterprises, enabling the collection, processing, and analysis of vast amounts of data from various sources. These tools empower decision-makers by providing insights that drive operational efficiency, revenue generation, and strategic decision-making. While BI tools primarily focus on querying and reporting business data, their capabilities extend to data visualization, ad hoc analysis, and real-time insights.

In recent years, the financial industry has witnessed significant transformations, particularly in trading practices. Physical trading floors are increasingly relying less on human traders, shifting towards algorithmic trading systems. The surge in data volume, coupled with advancements in technology, has reshaped securities trading and financial market architecture. Electronic trading has blurred boundaries between trading firms, accelerated competition, and heightened the need for transparency and speed of execution.

Algorithmic trading, characterized by automated trading using predefined rules, has emerged as a prominent innovation in the electronic trading arena. It offers benefits such as increased efficiency, transparency, and cost reduction in trade execution.

However, successful implementation requires a fully integrated low-latency infrastructure, robust risk management integration, and performance measurement capabilities. This paper takes an exploratory look at algorithmic trading from the perspective of technology personnel responsible for deploying and managing such infrastructure. Building upon background coverage of electronic and algorithmic trading, the paper utilizes grounded theory research methods to analyse interviews with senior financial community staff. The resulting theories address the adoption and implementation challenges of algorithmic trading systems, focusing on risk management, performance, and compliance.

In the context of this project, we aim to leverage the capabilities of BI tools to visualize critical information related to algorithmic trading systems. By integrating a selected open-source BI tool with a designed and implemented Data Warehouse (DW), we seek to empower decision-makers with actionable insights derived from algorithmic trading data. This endeavour underscores the importance of bridging the gap between raw data and meaningful insights, ultimately fostering informed decision-making in the dynamic landscape of algorithmic trading.

II. Project Overview

The project encompasses the development and integration of two key components: a Data Warehouse (DW) and a Business Intelligence (BI) tool. The DW serves as the central repository for storing large volumes of data generated by algorithmic trading systems, including market data, trade executions, and performance metrics. It is designed to handle complex data structures efficiently and ensure data integrity and consistency.

In parallel, the BI tool is selected and integrated to provide advanced visualization capabilities for the data stored in the DW. The BI tool enables users, including traders, analysts, and decision-makers, to interactively explore and analyse algorithmic trading data through intuitive dashboards, reports, and data visualizations. By leveraging the BI tool's capabilities, users can gain deeper insights into market trends, trading performance, and risk exposure, facilitating more informed decision-making processes.

The integration of the DW with the BI tool involves various stages, including data extraction, transformation, and loading (ETL), schema design, data modelling, and dashboard development. Throughout the project, attention is paid to scalability, performance optimization, and security considerations to ensure the reliability and usability of the integrated solution.

Key objectives of the project include:

1. Designing and implementing a scalable and efficient Data Warehouse architecture capable of handling large volumes of algorithmic trading data.
2. Selecting and integrating a suitable open-source or commercial BI tool that meets the visualization and analysis requirements of algorithmic trading systems.
3. Developing interactive dashboards, reports, and visualizations within the BI tool to present key insights derived from algorithmic trading data.
4. Ensuring seamless integration between the DW and BI tool, enabling real-time or near-real-time access to updated trading data.
5. Conducting thorough testing and validation of the integrated solution to verify its functionality, performance, and reliability.

Overall, the project aims to bridge the gap between raw algorithmic trading data and actionable insights through the development and integration of a robust Data Warehouse with a powerful Business Intelligence tool, ultimately empowering stakeholders to make informed decisions in the dynamic and competitive landscape of algorithmic trading.

III. UML Class Diagram[13]

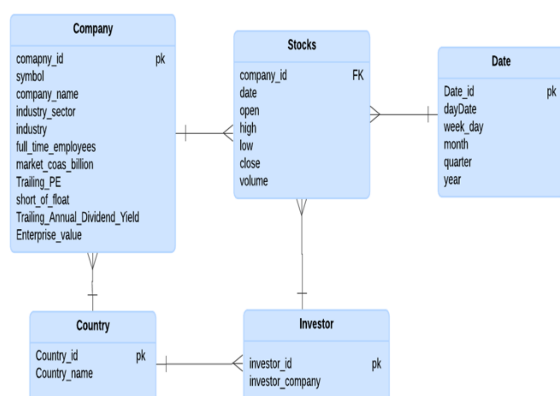


Figure: DW Dimension Model UML Class Diagram

IV. SCHEMA[10]

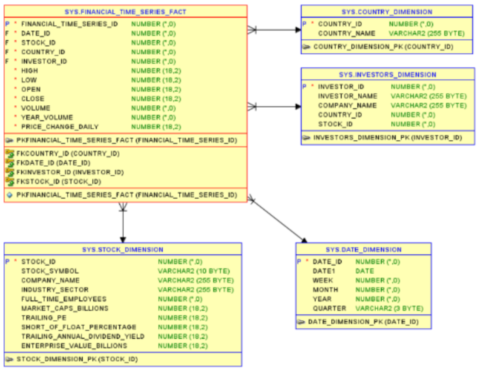


Figure: Star Schema for DW design

V. Dimensions

Oracle and SQL Developer to create dimensions for the Data Warehouse.

- "Stock_Dimension" contains essential information about the top 10 companies, including their stock symbols, market capitalization, trailing P/E ratio, and enterprise value.
- The individual stock data files contain time-series data for each company, including Date, Open, High, Low, Close, and Volume
- The "investor Dimension" contains details like invester_id and investor_company.
- The "Country Dimension" contains the Country names and their id.
- The "Date Dimension" contains time-series data including Date, Week, Quarter, and Year
- Using these dimensions, the FACT table is generated after completing the ETL process.

VI. ETL Pipeline

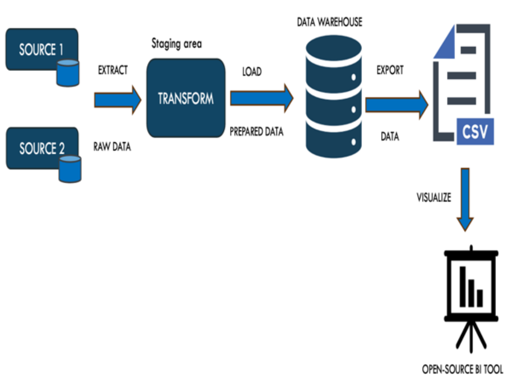


Figure: ETL Pipeline for DW

V. ETL Process

The initial extraction phase of the ETL process is fundamental in acquiring accurate and comprehensive data for subsequent analysis and decision-making in the stock market domain. This section introduces the objective and significance of the extraction phase.

1. Extraction

Data Sources Identification: emphasizes the importance of selecting reputable sources to ensure data quality and reliability.

Stock Market Data Extraction: elaborates on the methodologies used to extract historical and real-time stock market data from APIs or data feeds provided by reputable data providers. It outlines the essential stock attributes extracted, including open, high, low, and close prices (OHLC), trading volume, company information, and sector details.

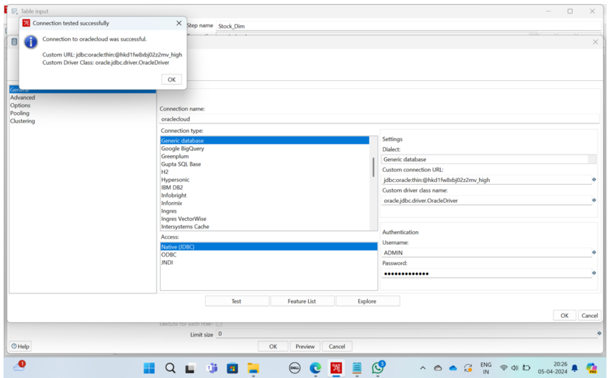
Country Information Extraction: the process of acquiring country-related data from public datasets, government sources, or third-party providers. It highlights the extracted country attributes such as names and regions.

Investor Information Extraction: the methodologies employed to retrieve investor-related data from financial institutions, regulatory bodies, or publicly available sources. It discusses the extracted investor attributes, including names, types, locations, portfolios, and trading activities.

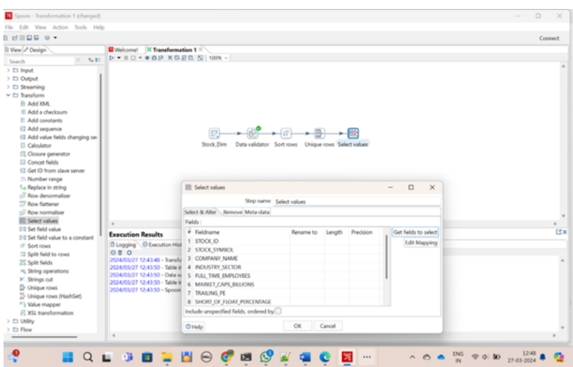
Date Dimension Generation: explains the creation of a date dimension table covering the desired date range. It discusses the inclusion of date-related attributes such as day, week, month, quarter, year, holidays, and other calendar-related information.

2. Transformation and Loading[15]

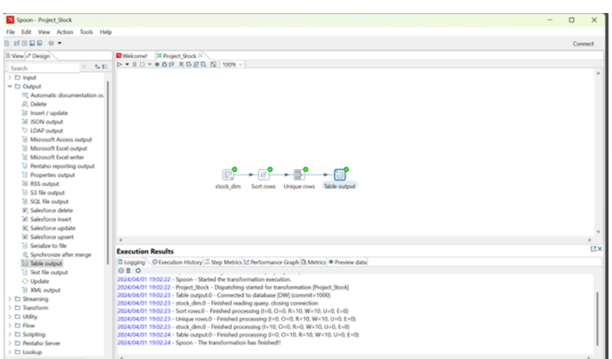
To perform the transformation and loading process Pentaho and Oracle Cloud has been connected.



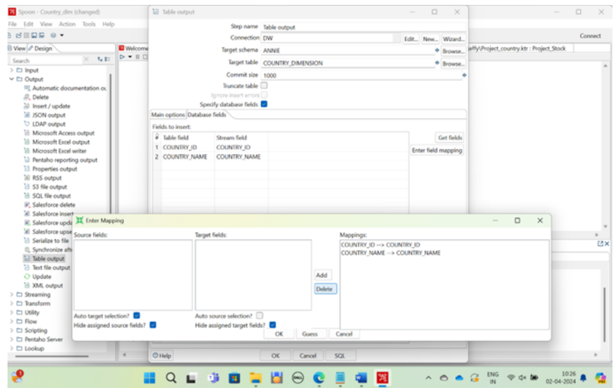
Transformation of Stock Dimension: During the transformation phase, raw stock data undergoes standardization and enrichment processes. This involves mapping stock ID to company stock ID, categorizing stocks, and rectifying any inconsistencies in the data.



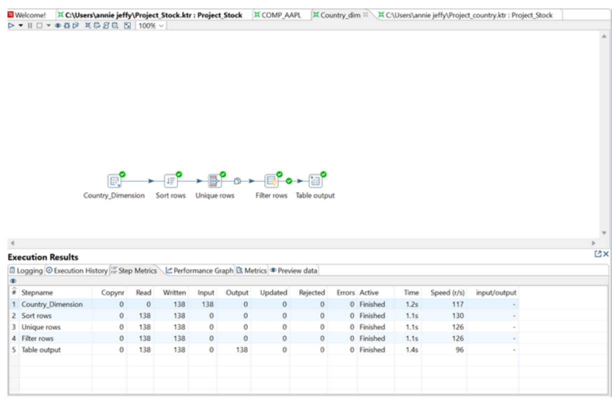
Loading of Stock Dimension: Transformed stock data is loaded into the data warehouse or database. This dimension table serves as a reference for analyzing stock performance, sector trends, and industry comparisons.



Transformation of Country Dimension: The transformation phase involves standardizing and enriching raw country data. This includes the data by regions and country id.

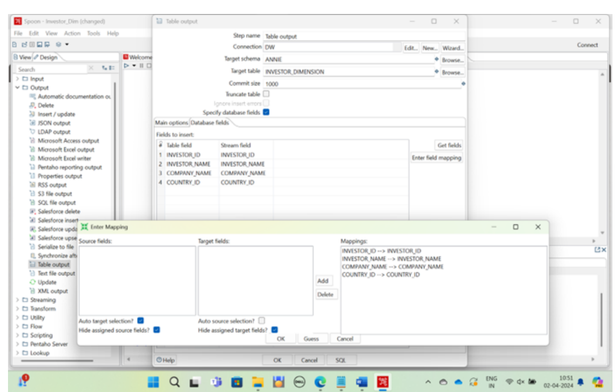


Loading of Country Dimension: Transformed country data is loaded into the data warehouse or database. This dimension table serves as a reference for analyzing global market trends.



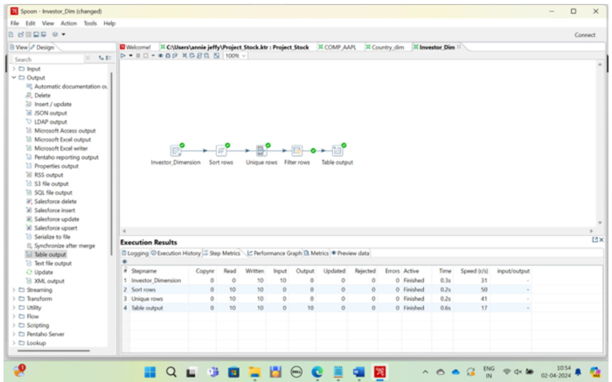
Transformation of Investor Dimension:

Raw investor data undergoes standardization and enrichment processes during transformation. This includes categorizing investors by type, aggregating portfolio holdings, and calculating performance metrics.



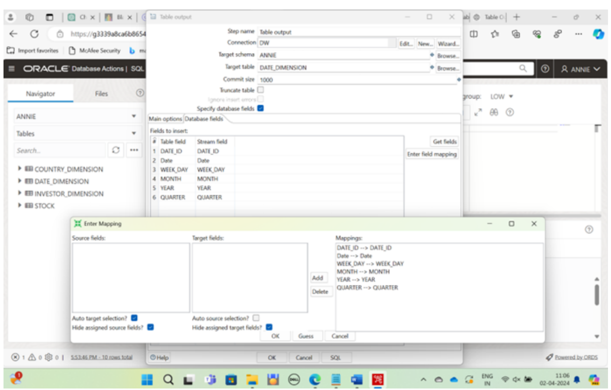
Loading of Investor Dimension:

Transformed investor data is loaded into the data warehouse or database. This dimension table serves as a reference for analyzing investor behavior, portfolio compositions, and market participation.



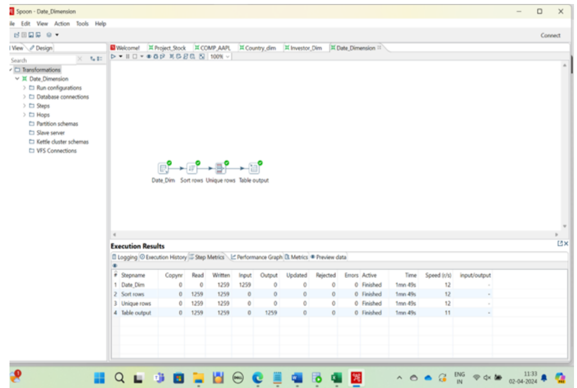
Transformation of Date Dimension:

The transformation phase involves generating a comprehensive calendar of dates covering the desired date range for analysis.



Loading of Date Dimension:

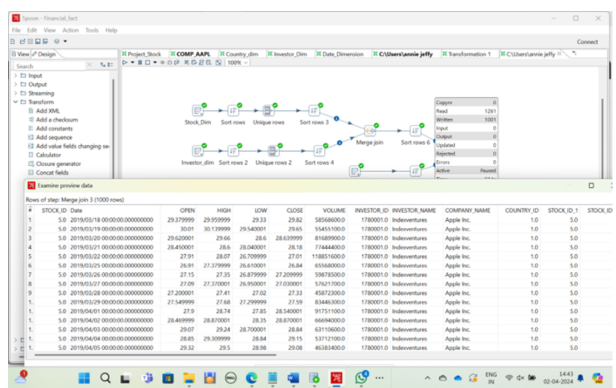
The date dimension is loaded directly into the data warehouse or database. This dimension table serves as a reference for analysing time-based trends, seasonality, and temporal patterns in the stock market data.



Transformation of Financial Fact Table:

During this phase, raw data extracted from various dimensions is transformed to make it suitable for loading into the fact table. This transformation process includes:

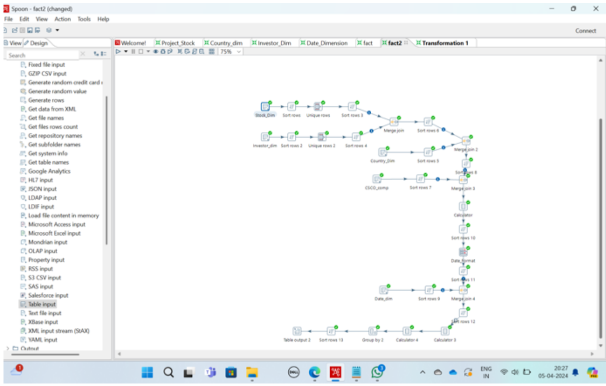
- Mapping dimension attributes to their corresponding IDs.
- Aggregating and summarizing data to derive key metrics such as daily volumes, high, low, open, and close prices, as well as total volume for each day.
- Ensuring data consistency and integrity by validating against predefined business rules and requirements.



Loading the Financial Fact Table:

Once the data is transformed, it is loaded into the Data Warehouse. This loading process involves:

Utilizing Pentaho(ETL tool) to load the transformed data into the fact table.



VII. Visualization

The primary objective of visualization is to provide users with a dynamic platform for exploring historical stock data, enabling them to compare the performance of different companies over time.

An interactive visualization tool has been built using Python. Data preprocessing for visualization involves reading the Excel files exported from DW, organizing the data into Pandas Data Frames, handling missing values (if any), and ensuring consistency in data format and structure.

Incorporating Plotly, a dynamic visualization library, our project presents a compelling line chart. Spanning from 2019 to 2024, this chart displays the closing prices of handpicked companies. Time is represented on the x-axis (Date), while stock closing prices are depicted on the y-axis. Each company's performance is illustrated by individual lines, facilitating effortless comparison.

The visualization goes beyond static displays with interactive features. Hovering over data points reveals comprehensive details such as company name, date, opening price, high, low, closing price, and volume. This interactivity enhances user engagement and facilitates deeper analysis.

Furthermore, the plot adapts dynamically based on user input, ensuring a seamless and intuitive experience. This feature empowers users to tailor their analysis effortlessly, making informed decisions effortlessly.

Seamless integration with ipywidgets[16] enhances user interactivity and exploration. By incorporating ipywidgets, we introduce a checkbox widget that empowers users to dynamically select or deselect companies. This checkbox widget conveniently displays a list of companies, enabling users to toggle their visibility on the plot with ease.

One of the key benefits of this integration is the instantaneous update of the plot upon selecting or deselecting companies. This feature allows users to observe real-time changes, enhancing their exploration experience.

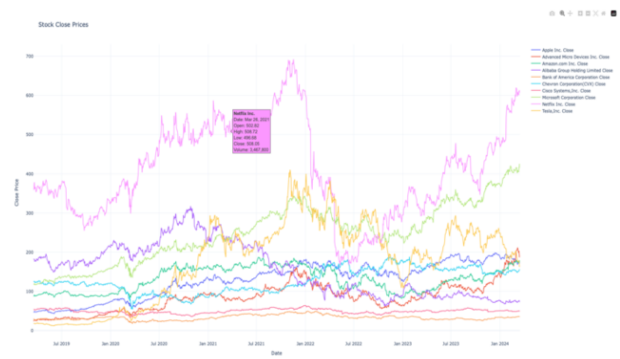


Figure: Top 10 company stocks from 2019- 2024

Figure: Comparison between stocks of 2 Companies: Apple vs Cisco

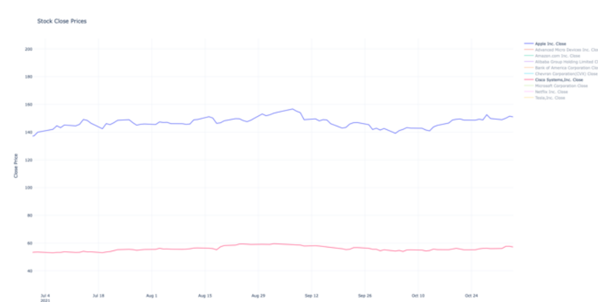


Figure: For Quarter

VIII. Conclusion

The project has successfully achieved its objectives of integrating a Data Warehouse (DW) with a Business Intelligence (BI) tool for algorithmic trading systems. The ETL process was efficiently implemented, ensuring the extraction, transformation, and loading of large volumes of trading data from diverse sources. By adopting a Star Schema design, the DW architecture provided a scalable and organized structure for storing and retrieving trading data, facilitating meaningful analysis. Moreover, the visualization aspect of the project was executed with precision, utilizing Python libraries such as Plotly to create interactive and insightful visualizations. The integration of ipywidgets[16] enhanced user interactivity, allowing stakeholders to dynamically explore data and make informed decisions in real-time. Overall, the project has bridged the gap between raw algorithmic trading data and actionable insights, empowering stakeholders with the tools and information necessary to navigate the dynamic landscape of financial markets. Through the seamless integration of DW with BI tools and advanced visualization techniques, the project has contributed to enhancing operational efficiency, identifying market trends, and mitigating risks in algorithmic trading systems.

IX. References

[1] youry/AlgorithmicTrading. (n.d.). GitHub. https://github.com/youry/AlgorithmicTrading

[2] Stock Market, Finance & Business News | Yahoo Finance Canada. (n.d.). Yahoo Finance. https://ca.finance.yahoo.com/

[3] Nasdaq. (n.d.). How advisors leverage technology to evolve and scale their businesses. Nasdaq. https://www.nasdaq.com/

[4] Li, C., Hains, G., Khmelevsky, Y., & Lee, W. (2012). Generating a Real-Time Algorithmic Trading System Prototype from Customized UML Models (a casestudy). ResearchGate.https://www.researchgate.net/publication/278629489_Generating_a_Real-Time_Algorithmic_Trading_System_Prototype_from_Customized_UML_Models_a_case_study

[5] Q. T. Minh, D. T. Thai, B. T. Duc, and N. H. Phat, "Designing a Data Warehouse Framework for Business Intelligence," published in the 2022 International Conference on Data Analytics for Business and Industry (ICDABI), Sakhir, Bahrain, 25-26 October 2022. DOI: 10.1109/ICDABI56818.2022.10041706

[6] Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). Data integration flows for business intelligence. In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (pp. 1–11). https://doi.org/10.1145/1516360.1516362

[7] T.A. Oketunji and R.O Omodara "Design of Data Warehouse and Business Intelligence System: A case study of a Retail Industry" [master's thesis]; Blekinge Institute of Technology; Karlskrona; Sweden; 2011. https://www.divaportal.org

[8] C. M. Păuna, "Automated Trading Software. Design and Integration in Business intelligence Systems.," ResearchGate, Jul. 2018, [Online]. Available: https://www.researchgate.net/publication/326508545_Automated_Trading_Software_Design_and_Integration_in_Business_Intelligence_Systems

[9] N. Ebadifard et al., "Data extraction, transformation, and loading process automation for algorithmic trading," arXiv, vol. 2312, pp. 12774–12774, Dec. 2023, [Online]. Available: https://arxiv.org/pdf/2312.12774.pdf

[10] "Data Mining: Concepts and Techniques: Han, Jiawei, Pei, Jian, Tong, Hanghang: 9780128117606

[11] "Yahoo is part of the Yahoo family of brands." https://ca.finance.yahoo.com/

[12] Plotly. (n.d.). https://plotly.com/graphing-libraries/

[13] Intelligent Diagramming | LucidChart. (n.d.). Lucidchart. https://lucid.app/lucidchart/825f0467-c6f1-4a95-b079-89530665584c/edit?invitationId=inv_e2937c6e-bb9a-4528-9cac-7a7514544e6c&page=0_0#

[14] Overview | Oracle Academy. (n.d.). https://academy.oracle.com/en/oa-web-overview.html

[15] Pentaho Data Integration and Analytics. (n.d.). https://www.hitachivantara.com/pentaho/pentaho-plus-platform/data-integration-analytics.html

[16] Jupyter Widgets — Jupyter Widgets 8.1.2 documentation. (n.d.). https://ipywidgets.readthedocs.io/en/stable/