Title: "Data Integration and Analysis with Talend: Empowering Insights from the 2023 Collapsed Banks Stock Prices Dataset ; Credit Suisse “

Done by :

* EMSSAAD Ilyass
* BENAABIDATE Ahmed Walid

Introduction:

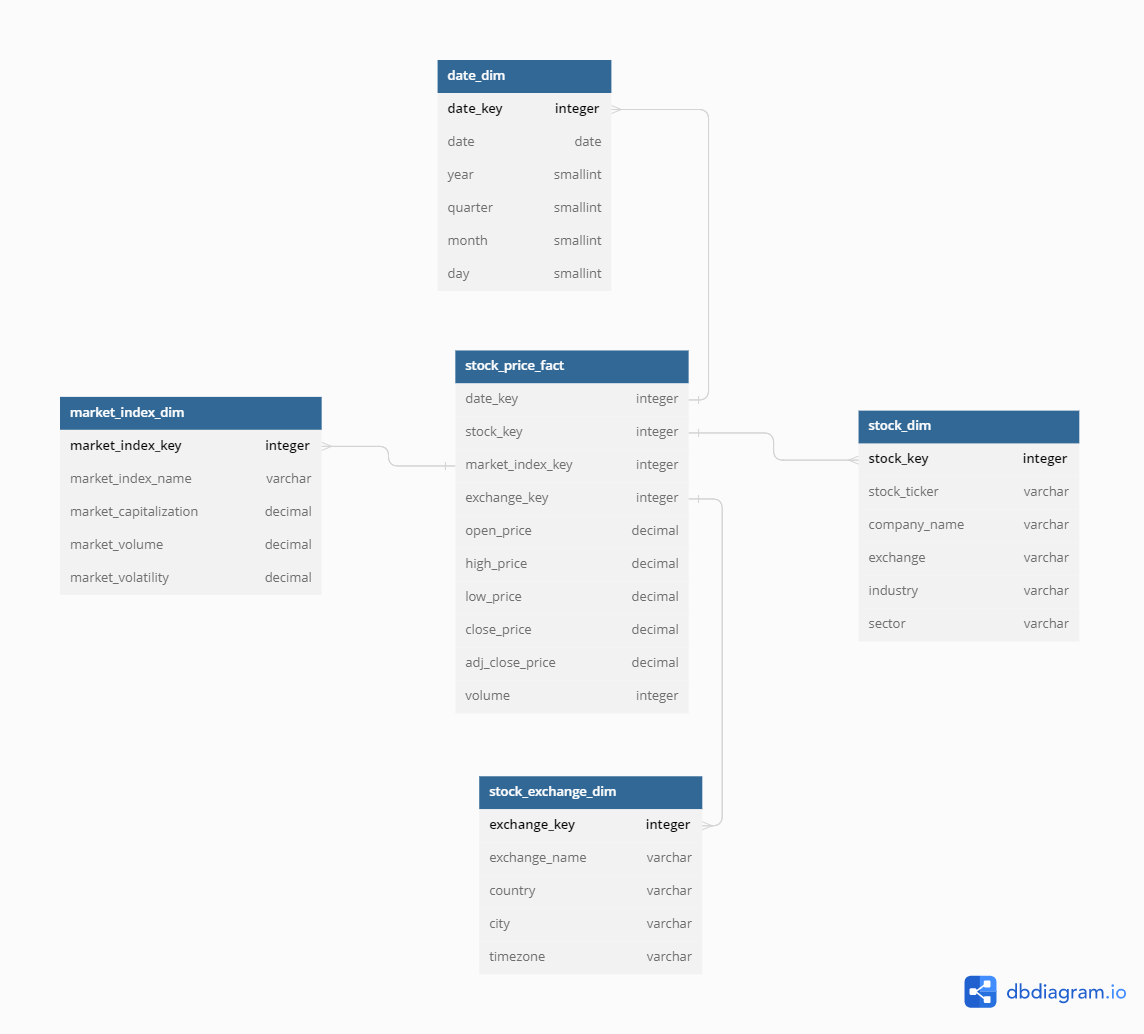
For our school project, we undertook a comprehensive business intelligence initiative using Talend, MySQL, and Power BI. Our objective was to analyze the dataset "2023 Collapsed Banks Stock Prices 2017-2023 Credit Suisse (CS.csv)" and derive meaningful insights. By harnessing the capabilities of Talend for data integration, MySQL for data storage, and Power BI for visualization, we aimed to showcase the end-to-end process of business intelligence.

Data Source and Tools:

In our project, we created a star schema to organize the data and facilitate analysis. The star schema is a type of database schema that consists of a central fact table connected to several dimension tables. The fact table contains the measures or metrics of interest, such as sales, revenue, or in this case, stock prices. The dimension tables contain the descriptive attributes or dimensions associated with the measures, such as date, stock ticker, company name, exchange, industry, and sector.

To create the star schema, we first designed the fact table, which contained the key metrics for the analysis. This included the date, open, high, low, close, adjusted close, and volume values for each stock. Next, we created dimension tables for each of the descriptive attributes associated with the measures. This included the date dimension, which contained the year, quarter, month, and day for each date; the stock dimension, which contained the stock ticker, company name, exchange, industry, and sector for each stock; the stock exchange dimension, which contained the exchange name, country, city, and timezone for each exchange; and the market index dimension, which contained the market index name, market capitalization, market volume, and market volatility.

By organizing the data into a star schema, we were able to simplify and optimize the analysis process. The fact table and dimension tables are linked together through keys, allowing for quick and easy data retrieval. This facilitated the creation of reports, dashboards, and visualizations, which provide valuable insights into the performance and trends of various stocks and markets.



The fact table in the star schema we created is called "stock\_price\_fact". This table contains the key measures for analysis, such as open price, high price, low price, close price, adjusted close price, and volume. The fact table has two foreign keys that link to the dimension tables: "date\_key" and "stock\_key". The "date\_key" column links to the "date\_dim" dimension table, which contains information about the date, including the year, quarter, month, and day. The "stock\_key" column links to the "stock\_dim" dimension table, which contains information about the stock, including the stock ticker, company name, exchange, industry, and sector.

Here are the dimension tables in our star schema:

1. Date Dimension Table: This table contains information about the dates for which stock data was collected. The columns in this table include date\_key, date, year, quarter, month, and day. The date\_key is the primary key for this table, which is used as a foreign key in the fact table. This table allows us to analyze stock data by date, year, quarter, month, and day.

2. Stock Dimension Table: This table contains information about the individual stocks being analyzed. The columns in this table include stock\_key, stock\_ticker, company\_name, exchange, industry, and sector. The stock\_key is the primary key for this table, which is used as a foreign key in the fact table. This table allows us to analyze stock data by stock ticker, company name, exchange, industry, and sector.

3. Stock Exchange Dimension Table: This table contains information about the exchanges where the stocks are traded. The columns in this table include exchange\_key, exchange\_name, country, city, and timezone. The exchange\_key is the primary key for this table, which is used as a foreign key in the stock dimension table. This table allows us to analyze stock data by exchange, country, city, and timezone.

4. Market Index Dimension Table: This table contains information about the market indexes being analyzed. The columns in this table include market\_index\_key, market\_index\_name, market\_capitalization, and market\_volume. The market\_index\_key is the primary key for this table, which is used as a foreign key in the fact table. This table allows us to analyze stock data by market index, market capitalization, and market volume.

A screenshot of a computer

Description automatically generated

By creating these dimension tables and linking them to the fact table using foreign keys, we can provide additional context for the key measures in the fact table, such as the stock ticker, company name, exchange, industry, sector, market capitalization, and market volume. This allows for more complex and nuanced analysis of the stock data, as we can now slice and dice the data along different dimensions.

In summary, the fact table serves as the central table in the star schema, containing the key metrics or measures for analysis, while also linking to the dimension tables through foreign keys. This organization allows for easy and efficient querying and analysis of the data.

A computer screen shot of a computer screen

Description automatically generated with low confidence

Our primary data source for the analysis was the "2023 Collapsed Banks Stock Prices 2017-2023 Credit Suisse (CS.csv)" dataset. This dataset provided a comprehensive collection of stock price data for Credit Suisse over the period from 2017 to 2023. It served as a valuable resource for conducting analysis and gaining insights into the stock market performance of Credit Suisse.

A screenshot of a computer

Description automatically generated

The dataset consisted of several columns that captured various aspects of the stock price data. Let's explore each column in detail:

1. Date:

The Date column represented the specific date associated with each stock price entry. It allowed us to track the temporal dimension and analyze how stock prices varied over time.

2. Open:

The Open column indicated the opening price of the stock for a particular trading day. It provided insights into the initial valuation of the stock at the beginning of a trading session.

3. High:

The High column represented the highest price reached by the stock during the trading day. It provided information on the stock's peak value within a given period.

4. Low:

The Low column captured the lowest price reached by the stock during the trading day. It provided insights into the stock's lowest valuation during a particular period.

5. Close:

The Close column indicated the closing price of the stock at the end of a trading day. It represented the final valuation of the stock for that particular day.

6. AdjClose:

The AdjClose column denoted the adjusted closing price of the stock. It took into account any corporate actions, such as stock splits or dividends, to provide a more accurate representation of the stock's performance.

7. Volume:

The Volume column represented the trading volume for a given day, indicating the number of shares traded during the trading session. It provided insights into the level of market activity and liquidity for the stock.

Using these columns, we were able to analyze various aspects of Credit Suisse's stock performance, including price trends, volatility, trading volume, and potential correlation with other market factors.

To leverage the data effectively, we utilized a combination of tools. Talend served as our data integration tool, facilitating the extraction, transformation, and loading of the dataset. MySQL, a relational database management system, provided a reliable platform for storing and managing the data. Lastly, Power BI enabled us to visualize and analyze the data through interactive reports and dashboards, offering a comprehensive view of Credit Suisse's stock market performance.

A picture containing diagram

Description automatically generated

By combining the rich dataset and powerful tools, we were able to gain valuable insights into the stock price dynamics of Credit Suisse and present our findings effectively.

ETL process:

In the ETL process for this project, the extraction step involved retrieving data from the "cs.csv" file and selecting specific columns for further processing. The columns present in the dataset are Date, Open, High, Low, Close, AdjClose, and Volume.

1. Extraction from "cs.csv":

We started by using Talend's tFileInputDelimited component to read the "cs.csv" file. This component allowed us to specify the file path and set the delimiter as appropriate for the CSV file. By configuring the component, we successfully extracted the entire dataset from the file.

Graphical user interface, application

Description automatically generated

2. Selecting Columns:

After the extraction, we focused on selecting specific columns for further analysis. In this case, we identified the following columns of interest from the dataset:

- Date: Represents the date of the stock price data.

- Open: Indicates the opening price of the stock.

- High: Represents the highest price reached by the stock.

- Low: Indicates the lowest price reached by the stock.

- Close: Represents the closing price of the stock.

- Volume: Indicates the trading volume for the given date.

By narrowing down our focus to these specific columns, we ensured that our subsequent data processing steps and analysis would revolve around these key attributes.

The extraction process involved reading the "cs.csv" file and selecting the desired columns for further analysis and processing. This step set the foundation for the subsequent steps in the ETL process, where we performed data transformation, loading, and analysis using tools like Talend, MySQL, and Power BI.

Data Transformation:

To begin, we imported the "2023 Collapsed Banks Stock Prices 2017-2023 Credit Suisse (CS.csv)" dataset into Talend. We performed data integration tasks, such as examining the data quality, handling missing values, and addressing duplicates. Talend's intuitive interface enabled us to navigate through the data transformation process seamlessly. We extracted relevant attributes, including stock prices, dates, and additional pertinent information, while ensuring data consistency and integrity.

Transformation Steps:

1. Filtering Data based on Volume:

The first step in the data transformation process was to filter the dataset based on the volume column. We applied a condition to retain only the rows where the volume was greater than 0. This filter ensured that we focused on relevant data points with actual trading activity.

A screenshot of a computer

Description automatically generated with medium confidence

2. Handling Duplicates and Unique Records:

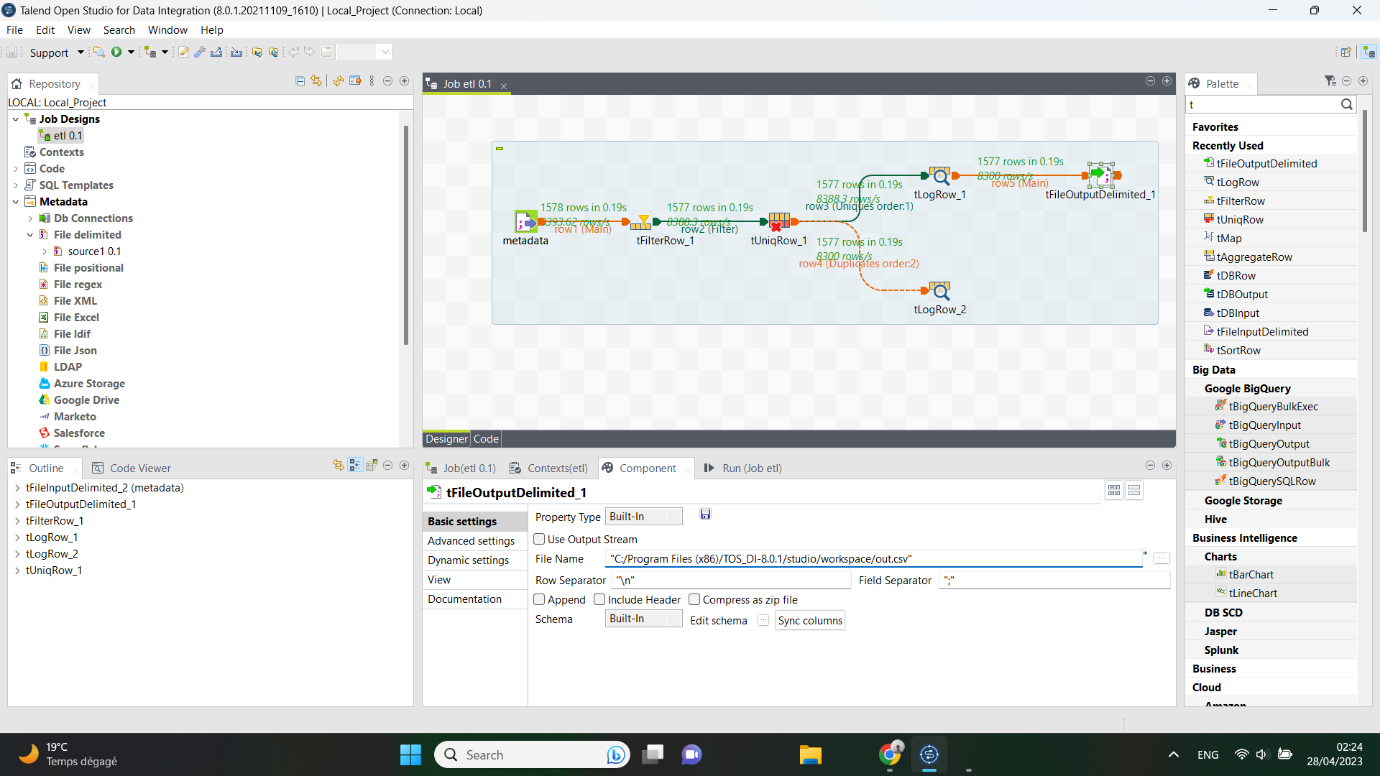
After filtering the data, we proceeded to handle duplicates and unique records. We used the tFilterRow component in Talend to separate the dataset into two streams: one containing duplicate records and the other containing unique records. This step allowed us to identify and isolate any duplicated entries within the dataset.

Graphical user interface

Description automatically generated

3. Generating CSV Output:

For the unique records stream, we utilized the tFileOutputDelimited component. This component allowed us to specify the output format as CSV (Comma-Separated Values). We configured the component to write the filtered unique records to a CSV file. This generated a separate file containing only the unique records meeting the volume criteria.



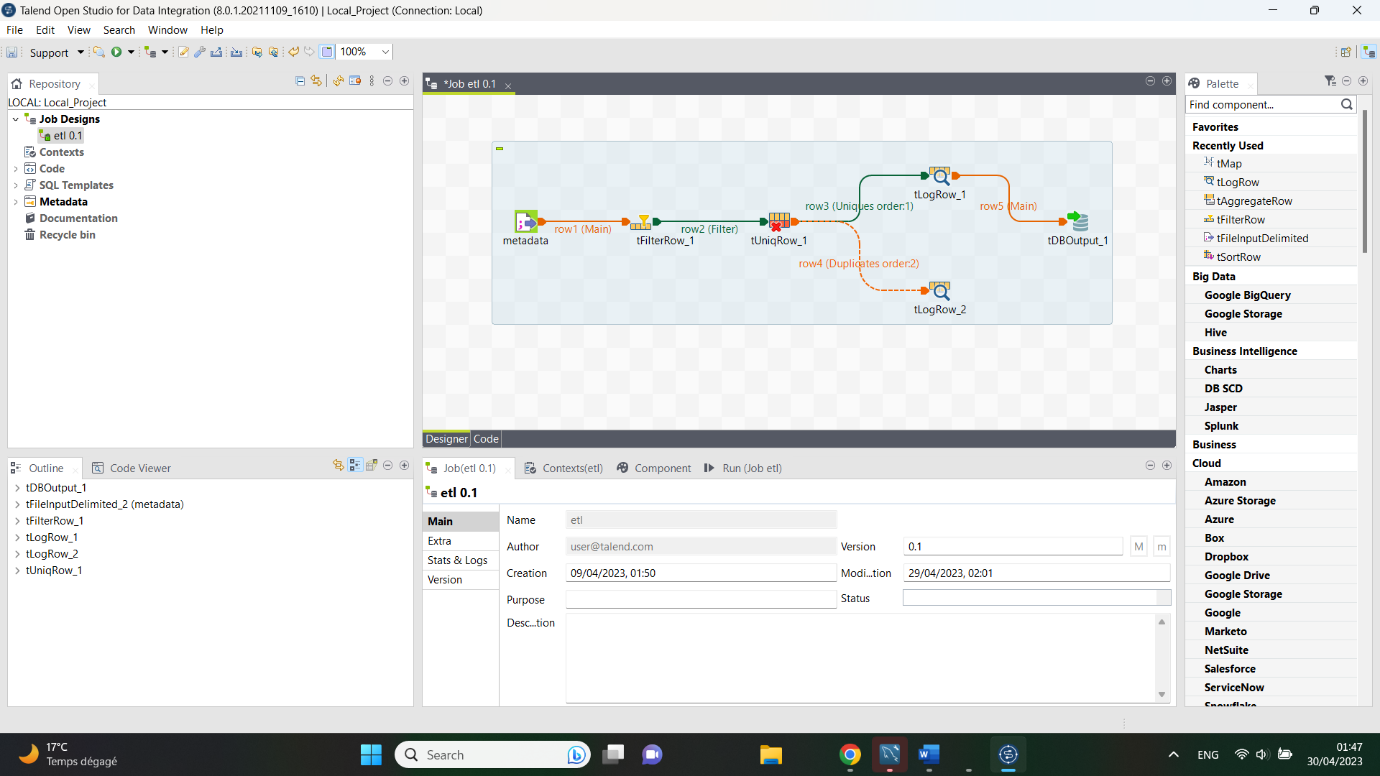
A screenshot of a computer

Description automatically generated with medium confidence

Data Loading into MySQL:

To store the transformed dataset, we employed MySQL as our database. Using Talend's MySQL connectors, we established a connection to the database. We designed a suitable table structure within MySQL, ensuring appropriate column types to accommodate the dataset's attributes. Talend facilitated the loading of the transformed data into MySQL, guaranteeing data integrity and reliability for subsequent analysis.

To load the filtered and unique records into MySQL, we employed the tMySQLOutput component. This component established a connection to the MySQL database, allowing us to define the table structure and map the columns of the dataset to the corresponding table columns in MySQL. We configured the component to insert the unique records directly into the MySQL table.



By following these transformation steps, we filtered the dataset based on volume, separated duplicate and unique records, generated a CSV file containing the unique records, and loaded them into the MySQL database for further analysis or reporting purposes.

Graphical user interface, text, application

Description automatically generated

Data Analysis and Visualization with Power BI:

With the dataset successfully loaded into MySQL, we leveraged Power BI's advanced visualization capabilities to gain insights from the data. We connected Power BI to the MySQL database, accessing the transformed dataset. In Power BI, we performed various data analysis tasks to explore different aspects of the dataset. This involved analyzing stock price trends, identifying daily fluctuations, and evaluating trading volumes. We employed a range of visualizations, including line charts, bar graphs, and heat maps, to present the analyzed data effectively.

Graphical user interface, application

Description automatically generated

The Swiss bank's share price since the start of 2020, and the key events that marked its decline

A picture containing chart

Description automatically generated

Summary of Credit Suisse's Challenges and Events:

Credit Suisse, a prominent Swiss bank, has faced numerous challenges and events that have significantly impacted its operations and reputation. Here is a summary of some key incidents:

1. February 2020: Spying Scandal and CEO's Departure:

- Credit Suisse hired private detectives to spy on its former head of wealth management who joined a rival bank.

- CEO Tidjane Thiam's tenure ended abruptly, and the bank was criticized for misleading regulators.

2. March 2021: Collapse of Greensill Capital:

- Credit Suisse had substantial investments in Greensill Capital, a troubled financial firm specializing in short-term corporate loans.

- The collapse led to the closure of four funds and significant financial losses for Credit Suisse.

3. March 2021: Archegos Defaults:

- Credit Suisse suffered losses of $5.5 billion when the hedge fund Archegos Capital Management defaulted on its highly leveraged bets.

- Weak management and control failures were identified as contributing factors to the losses.

4. October 2021: Fine over Mozambique Scandal:

- Credit Suisse was fined $475 million by US and British authorities for its involvement in a bribery scandal related to loans in Mozambique.

- The bank's actions led to an economic crisis in Mozambique after the International Monetary Fund withdrew support.

5. January 2022: Chairman's Resignation:

- Former Lloyds Banking Group CEO Antonio Horta-Osorio, brought in to lead a turnaround, resigned after accusations of breaking Covid restrictions.

- He described the crisis at Credit Suisse as one of the worst he had experienced in his career.

6. February 2022: Suisse Secrets Data Leak:

- Credit Suisse faced allegations that it helped individuals involved in serious crimes to hide funds.

- A global media investigation revealed extensive data on bank accounts and alleged involvement in money laundering and corruption.

7. March 2022: Bermuda Judge's Ruling:

- Credit Suisse's local life insurance arm was ordered to pay damages of $553 million due to failures related to a fraud committed by a former banker.

8. June 2022: Money Laundering Conviction:

- Credit Suisse and a former employee were found guilty of failing to prevent money laundering by a Bulgarian cocaine-trafficking ring.

- The bank received a fine, and the former employee revealed that the bank continued managing funds despite knowledge of criminal activities.

9. October 2022: Turnaround Efforts and Outflows:

- New leadership announced a plan to cut jobs, raise capital, and focus on Credit Suisse's Swiss roots.

- However, the plan failed to impress investors, and significant customer outflows and financial losses were reported.

10. March 2023: Terminal Crisis and Acquisition by UBS:

- Credit Suisse faced SEC inquiries and admitted to financial control weaknesses, leading to a sharp decline in its shares.

- The Swiss National Bank facilitated a takeover by UBS, resulting in the end of Credit Suisse's 167-year history.

These events highlight the multiple challenges Credit Suisse faced, including governance failures, financial losses, regulatory issues, and reputational damage. The bank's future now lies under the umbrella of UBS as it undergoes a significant restructuring process.

Conclusion:

Through this project, we demonstrated the end-to-end process of business intelligence using Talend, MySQL, and Power BI. We showcased the importance of data integration, transformation, storage, analysis, and visualization in extracting valuable insights from raw data. By harnessing Talend's data integration capabilities, we successfully transformed and loaded the "2023 Collapsed Banks Stock Prices 2017-2023 Credit Suisse (CS.csv)" dataset into MySQL. Using Power BI, we visualized the data, uncovering actionable insights. This project highlighted the power of business intelligence in facilitating informed decision-making and demonstrated the practical applications of Talend, MySQL, and Power BI in real-world scenarios.