IS3107 Data Engineering

Academic Year 2024/2025 Semester 1

Course Project

Name	Student ID
Sohn Jonghyun	

1.1. Background of ESG Data Management

With the recent interest in responsible and sustainable investing, Environmental, Social, and Governance Data, also referred to as ESG data, has become an indispensable constituent in investment analysis. Investors have come to view the ESG metrics as an important non-financial measure when assessing the operations of a given company. This can either be in the form of a good return on investment that will take a long time to realize or aim at changing the way investment is done to fit an ethical standpoint and its sustainability. These ESG metrics are usually data points of carbon emissions, diversity, and inclusion, corporate governance and compliance, and the environment.

The rising consciousness of climate change, social equity and the importance of running a business ethically has led to a growing demand for ESG data. However, there are still several obstacles that are imposed in the management and use of ESG data. The problem of different data sources, varying reporting standards, and the need for data incorporation still exists and needs urgent attention. Consequently, to effectively manage and integrate different sources of ESG data to make better investment decisions, it is essential to have a sound and holistic data warehousing infrastructure.

1.2. Objective

This project aims to develop an ESG data warehouse tailored to the needs of buy-side investors. This data warehouse will both be scalable and comprehensive with a compilation of multiple companies' ESG data to provide a more cohesive warehouse system that can support various situations for analysis. This high-quality compilation of ESG data will facilitate investment decisions by providing insights into three different performance metrics: environmental, social, and governance.

By gathering data from reliable sources and constructing a suitable pipeline for data extraction, transformation, and loading (ETL), these procedures will simplify the process of data analysis and provide insights for investment actions. Further discussion on unique features of the data warehouse that can handle complex queries related to ESG metrics will be followed subsequently. throughout this report.

1.3. Importance of ESG Data for Investors

In the modern world, ESG data are widely used to assess long-term risks and opportunities associated with investors' portfolios. In terms of rating, higher ESG ratings suggest stronger corporate governance, better environmental practices, and more equitable social policies. Now all of these statistics determine the overall sustainability and financial health of the company.

It is also possible to utilize ESG data for many scenarios. Mitigating risk would be the most significant and top priority for all investors. With hands-on ESG data, investors can avoid potential future liabilities, reputational damage, and operational disruptions. On top of that, stronger ESG-profiled companies are likely to perform better long term and result in a successful return.

Incorporating ESG data within investment strategies is also beneficial for the investors in meeting the regulatory requirements but moreover addresses a fast-emerging market that is concerned with sustainability. Hence the analytical processing and dissemination of relevant ESG data will be crucial in providing the buy-side investor support geared at investment decision making.

2. Client Needs and Scope of Data

With the scope of this project being targeted towards buy-side investors who would be willing to purchase the project's data warehouse, these investors will likely use this data warehouse for further investment use regarding ESG attributes. Therefore, it will be essential to consider some of the queries that will be used to evaluate the companies. This can be split into a total of three sectors: environmental, social, and governance. The queries are categorized into relevant areas of analysis and will help lead the design and functionality of the data warehouse. This will, in turn, ensure a wide range support of for ESG-focused analyses.

2.1. Environmental Performance Queries:

1. Find the average ESG rating per year for each industry.
2. Find the total carbon emissions reported and the main environmental concern per year, for each company.
3. Find the average energy consumption per year for companies in a specific region.
4. Find the companies with the highest default rates in meeting environmental targets (e.g., emissions reduction goals) per year.

2.2. Social Responsibility Queries:

1. Find the overall employee turnover rate per year per company.
2. Analyze the impact of a company's diversity and inclusion (D&I) practices on its employee satisfaction scores.
3. Identify which companies report the most significant year-on-year improvement in their ESG ratings.

2.3. Governance Practices Queries:

1. Check if a company has records of governance violations (e.g., regulatory fines or scandals).
2. Analyze the relationship between corporate governance ratings and a company's financial performance (e.g., stock price or revenue growth).
3. Analyze the impact of a company's executive compensation structure on its overall governance rating.

2.4. Data Scope

The core focus of the dataset for the ESG data warehouse will constitute ESG Risk Ratings which provide information about the company's exposure to material ESG issues and its ability to handle risks. Additionally, it is a great tool for buy-side investors to assess a company's sustainability, corporate responsibility, and governance practices. Now, these ratings are split into a total of 5 categories: negligible, low, medium, high, and severe as displayed in Figure 1 below.

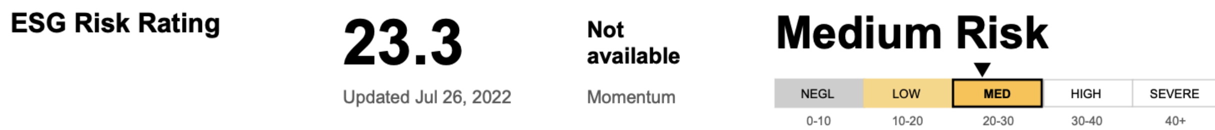

Figure 2.1 - Sustainalytics ESG Risk Rating Figure

These ratings offer a clear framework for comparing companies across both industry-level and global levels**.** Additionally, by integrating industry-specific material ESG risks, the ESG risk rating provides insights into how well companies adapt to challenges. For example, the company mentioned in Figure 2.1 having a risk rating of 23.3, can be viewed as a potential liability due to poor risk management. Also, it is easily trackable as ESG risk ratings are captured on year-over-year trends to highlight improvements or declines in ESG performance. Therefore, ESG risk rating already allows investors to analyze historical data for long-term trends and help detect early warning signs of risk escalation or recognize companies with stronger sustainability practices.

Nevertheless, solely relying on ESG risk rating to construct the data warehouse would not be sufficient enough. These ratings do provide limited context as it does not provide the raw data or detailed metrics (e.g. carbon emissions, employee turnover rates) that investors might be willing to key into for further analysis. On top of this, despite its captured trends being year-over-year, the scoring is quite static. The ratings are not updated periodically and this may not reflect real-time developments or nuanced shifts within a company's ESG strategies.

For such reason, in addition to ESG ratings, additional data that signifies environmental performance, social responsibility, and governance practices will be required to be reflected in the data warehouse. This will further enhance the ESG dataset by addressing gaps in ESG risk ratings and allowing it to be more comprehensive and actionable. The utilisation and a good mix of both commercial and free data sources will be worthy enough to be utilized to construct and bring unique values to the data warehouse. As for commercial data sources, these data provide more in-depth, validated, and regularly updated datasets. Especially, websites like Sustainalytics and MSCI ESG Ratings offer myriad amounts of data on carbon emissions, controversies research, and corporate governance indicators on top of ESG risk ratings which are tailored for investment insights. This ensures data quality and extensivity of the data but comes with the cost of payment. Free data sources, which can come from the World Resources Institute (WRI) or Carbon Disclosure Project (CDP), provide supplementary information that can support the details earned from commercial datasets. While these sources have limited extensivity and may be outdated, it is cost-effective (usually free of charge) and a great supporting tool in developing a robust ESG data warehouse. Therefore, a mix of these sources will foster the development of a well-constructed ESG data warehouse that is tailored for investors to carry on appropriate investment decision-making for various industries.

3. Data Pipelines

A more suitable data pipeline for this ESG data warehouse will be to take the ETL (Extract-Transform-Load) approach compared to the ELT (Extract-Load-Transform approach). The reason for this is firstly due to ETL's effectiveness in data cleaning and quality control. ESG and environmental data often come from various sources which might contain inconsistent, incomplete, or errors in each data. With ETL's structure allowing data cleaning and validation during the transformation step before loading into the warehouse, it ensures only high-quality and valid data goes into the final warehouse. In addition, most ESG data requires standardisation (e.g. aligning carbon emission units, normalizing dates), aggregation (e.g. creating industry benchmark), and enrichment (e.g. re-calculating ESG scores). Once again, the structure for ETL is more suitable in such conditions as there will be complex transformations required for the data to be unified and will be more cost-efficient to do such a process before all the ESG-related data are loaded into the warehouse. There is also a matter of resource efficiency as ELT heavily relies on the processing power of the warehouse to transform data past-loading. As there are myriad numbers of environmental metrics, this can strain the warehouse and make it less cost-effective. However, ETL completes such a procedure before the loading stage to allow the warehouse to focus on the querying and analytics aspects, improving its overall performance and cost-effectiveness. For the scenario of an ESG data pipeline, lambda architecture will be more suitable. The primary reason is for dual needs for batch and real-time processing. ESG data will require batch processing for historical reports and stream processing for real-time updates. Also, each layer can play a significant role in each request. The batch layer can provide accurate and fault-tolerant computations for long-term ESG performance evaluation while the speed layer can provide insights for stakeholders instantly with real-time dashboards. Compared to this, Kappa architecture's batch-like processing for processing many historical datasets and large streams of data might be not as effective.

3.1. Environmental Data Pipeline

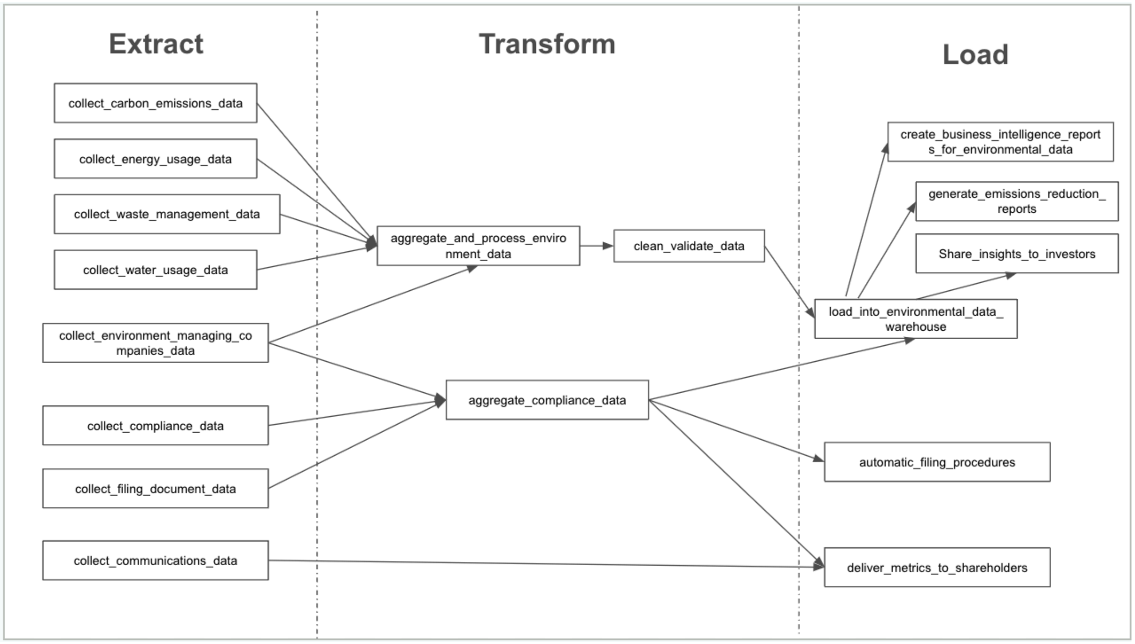

Figure 3.1 - Environmental Data Pipeline Figure

Data Sources Required: Carbon Emission Reports, Water & Energy Usage Records, Waste Management Data, Regulatory Documents, Environmental Requirements, External Environmental Databases.

Extracting Phase: Collect raw data from multiple sources related to environmental analysis such as carbon emissions disclosures, energy usage reports, waste data from companies, real-time data from sensors tracking energy consumption and so on.

Transformation Phase: Normalize units of measurement for consistency (kWh for energy) and align with industry standards and benchmarks for comparison. Also, data cleaning is carried out by handling missing values, removing duplicates, and addressing outliers in certain metrics. It is also important to validate data accuracy against external benchmarks and regulatory standards.

Loading Phase: Load all the data into the environmental data warehouse and organise processed data in a structured schema categorised into industry, company, time, and environmental factors. There could be additional tables that separate emissions, energy, waste, and compliance metrics. The creation of dashboards with key environmental metrics for real-time visualisation will be necessary for shareholders and investors to view current data streams.

3.2. Social Data Pipeline

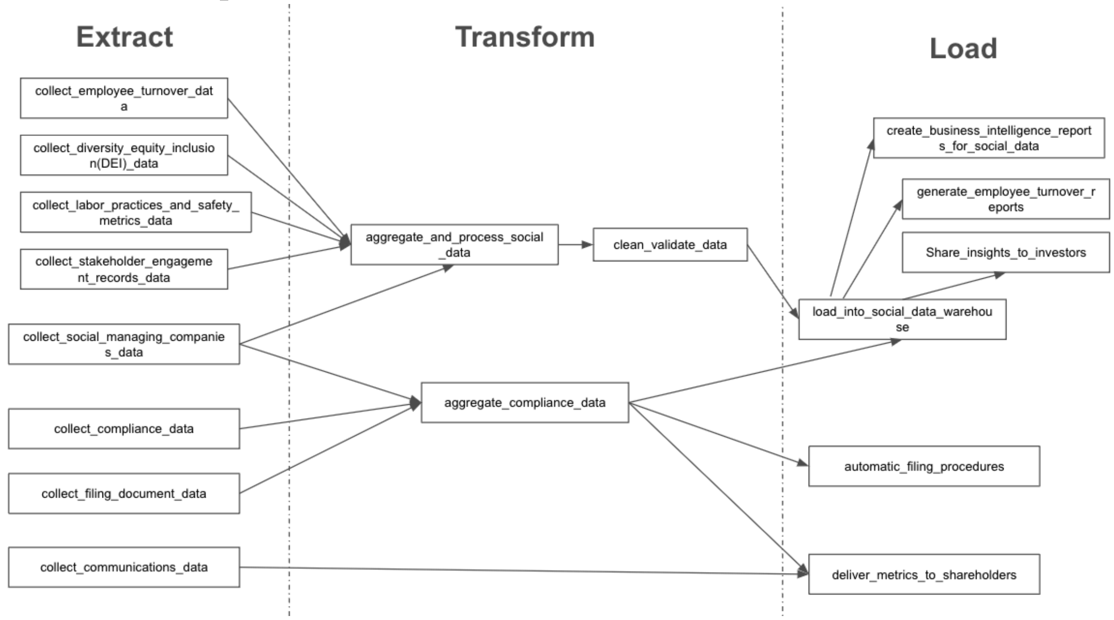

Figure 3.2 - Social Data Pipeline Figure

Data Sources Required: Employee turnover data, Diversity, Equity, and Inclusion (DEI) reports, Labor practices and safety metrics, stakeholder engagement records, social ratings and benchmarks.

Extracting Phase: Collect data by gathering social metrics from internal HR systems or external benchmarks and ratings from various data-providing platforms. Can also stream data on safety incidents or labour law violations from IoT-enabled monitoring systems.

Transformation Phase: Normalize units of measurement for consistency (turnover percentages, pay equity ratios) and align with industry standards and benchmarks for comparison. Also carry on data cleaning through handling missing values, removing duplicates, and addressing outliers in certain metrics (e.g. employee satisfaction or turnover ratio). It is also important to aggregate data at the company, regional, and industry levels.

Loading Phase: Load all the data into the environmental data warehouse and organise processed data in a structured schema categorised into demographics, and time. There could be additional tables that separate turnover, DEI, and labour practices metrics. The creation of dashboards with key social metrics for real-time visualisation will be necessary for shareholders and investors to view current data streams. Some possible creation of data can be turnover trends by department or region. Also, creating automated reports such as compliance reports to handle trends such as employment engagement will be a viable option.

3.3. Governance Data Pipeline

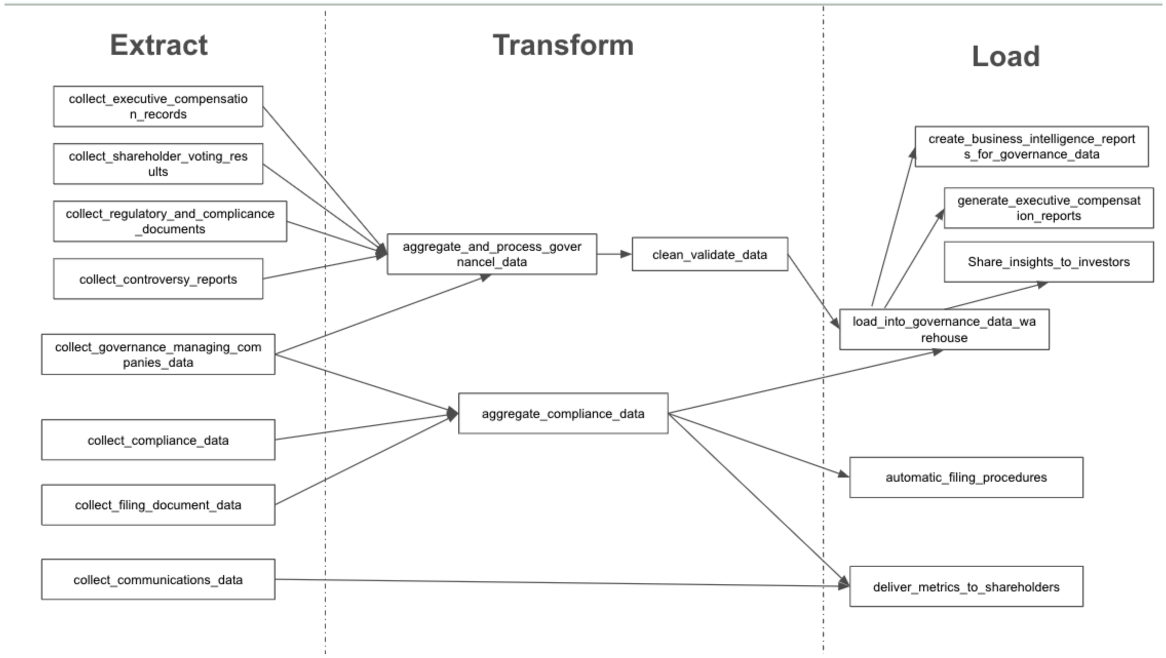

Figure 3.3 - Governance Data Pipeline Figure

Data Sources Required: Executive compensation records, Shareholder voting results, Regulatory and compliance documents, Controversy reports, Governance ratings and benchmarks

Extracting Phase: Collect data by gathering social metrics from corporate proxy filings and disclosures for board composition details (e.g. diversity, tenure, independence). Also, collect information from annual reports or public filings for shareholder voting outcomes. It is also possible to refer to ESG platforms such as Sustainalytics for more precise ratings and controversial data.

Transformation Phase: Normalize units of measurement for consistency (board diversity metrics (% of men or minorities)) and align with industry standards and benchmarks for comparison. Data cleaning is also carried out by handling missing values, removing duplicates, and addressing outliers in certain metrics. It is also important to aggregate data at the company, regional, and industry level including metrics such as pay-performance ratios and board independence ratios.

Loading Phase: Load all the data into the environmental data warehouse and organise processed data in a structured schema categorised into board composition, executive compensation, shareholder engagement, and compliance and controversy. The creation of dashboards with key social metrics for real-time visualisation will be necessary for shareholders and investors to view current data streams. Some possible creation of data can be board diversity trends or executive compensation comparisons. Also, creating automated reports such as compliance reports to adhere to governance standards will be necessary.

3.4. Validity and Difficulty in Implementation

The above-mentioned data pipeline is structurally valid and is easy to implement. The reason for this is justified through two main reasons: the ETL approach and the Lambada architecture. The ETL approach ensures a logical end efficient flow of data from raw sources, which are very common for ESG-related data, to actionable insights, effectively organized into fields the data is most suitable in. Furthermore, the pipeline also supports both real-time and batch processing making it align with the suitability of taking the Lambda approach. This approach allows both historical batch processing, such as analysis of year-over-year trends, and real-time updates that can be brought from stakeholder feedback systems. Thus, such flexibility in both processing and a well-thought-out connection with the ETL approach allows the pipeline most suitable for ESG reporting requirements and makes it moderately less difficult to implement.

3.5. Efforts for Improvements

There have been multiple efforts in trying to make improvements from the original plan to address ESG data challenges. One of the main focuses, as mentioned earlier, is data cleaning and validation. With a lot of ESG data having missing fields, outliers, and duplicates, improvements have been made within the ETL approach to make sure these outliers in the data are handled before loading and validated. Additionally, standardisation and aggregation have also been another focus while constructing the data pipeline. Some data will utilize different metrics from other reports which causes inconsistencies that will be required to be addressed. Such cases usually happen for metrics for CO2 emissions and turnover ratios which are all valuable key metrics in constructing the warehouse. Thus, with such inconsistencies handled, it simplifies the cross-company and industry-level comparisons. On top of that, real-time dashboards were incorporated into the design to cater for the needs of stakeholders' requests to access live insight and will ensure stakeholder's timely decision-making.

For improvements in scalability and query performance, categorizing the data warehouse into specific fields (e.g. emissions, DEI, governance metrics) improves the scaling of new data sources by enabling the independent processing of each field. Such an approach allows easier integration of new data sources and allows the constructed warehouse to scale efficiently.

There have been efforts made to address critical engineering goals. Data quality is ensured through the validation process, reducing resource strain and optimizing performance. The warehouse is more cost-effective as it focuses specifically on querying and analytics after it is loaded with processed data instead of managing computationally expensive transformation after the loading phase.

4. Decided data formats and DB software

The chosen data formats have been decided among various types of common ESG data types. The most considered two types are JSON and CSV which has great compatibility and the ability to handle both structured and semi-structured data, allowing seamless data extraction and transformation during the ETL process.

JSON - JSON is a type of format that is commonly used for semi-structured data which aligns with the ESG data types that are sometimes not captured or formatted in conventional ways (e.g. raw data for CO2 pollution). Therefore, its hierarchical structure allows easier representation of data that has a complex relationship with other data.

CSV - CSV is a lot more suitable for structure data and has a simple structure with lightweight capacity and compatibility. With such benefits, it holds the most benefit in working with the ETL structure and helps manage ESG-related structural data, which is also very common.

Therefore, as both data formats have their usability and limitations, having a mix of both data formats for conditions they are tailored to will maximise the efficiency overall. It would be most efficient to start with CSV for storing and querying processed data in the warehouse and utilize JSON for ingesting raw or sem-structured data.

As for the suitable DB software, two main considerations are present: MySQL and MongoDB. Both programs have their strengths and weaknesses.

MySQL - MySQL is best for structured data like emission records or governance benchmarks. It also supports complex joins, aggregations, and analytical queries, making it an optimal option for ESG reporting. It is also suitable for medium-sized datasets but less effective for large-scale horizontal scaling.

MongoDB - MongoDB is more suitable for semi-structured data such as nested governance reports. It is especially efficient for gathering nested or hierarchical data but less suitable for complex relational queries, which is very important for ESG data gathering. It does provide efficiency in managing large volumes or high-velocity real-time data but might not be applicable for this project as ESG ratings usually are announced year-to-year basis. Additionally, it is best for ingesting raw, semi-structured data before transformation.

Thus, despite MongoDB having multiple benefits in managing large volumes of real-time data, it will not be a suitable option for this project as the project's main focus is on ESG risk rating with additional information to back up the details for the rating. For that reason, MySQL will be more suitable for its ability to handle structured relational data, support analytics, and provide compatibility with Business Intelligence(BI) tools.

5. Data Warehouse

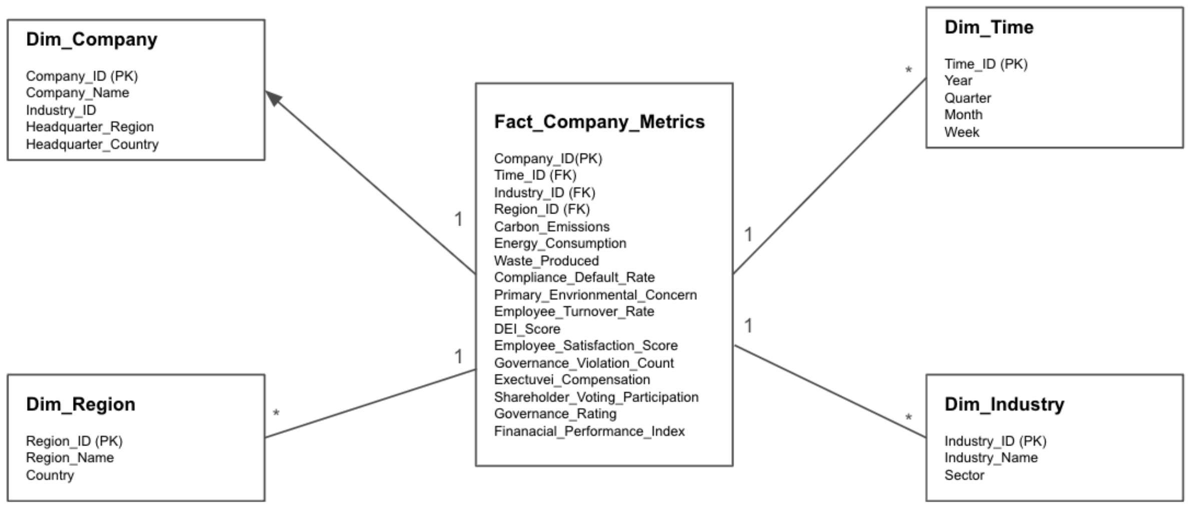

Figure 5.1 - Data Warehouse Figure

5.1. Design Approach

The initial step was to identify the business processes. The main idea is that ESG data is essential for analyzing a company’s performance across various fields such as environmental, social, and governance dimensions. Since the company is the main entity and evaluated various ESG metrics at the company’s level, it was most logical to set the primary key as the company’s ID.

This ESG data warehouse will employ the Kimball Dimensional Modeling approach which is most suitable for cases where some redundancy is tolerated and has an easier design for other business roles. Additionally, the design uses a star schema with one consolidated fact table (Fact_Company_Metrics) and four distinct dimension tables (Din_Time, Dim_Industry, Dim_Company, and Dim_Region). Such design allows simplicity, scalability, and performance in handling ESG-related data queries.

To further discuss each table component, instead of having multiple separate fact tables for the environmental, social, and governance portion, it would be a simpler approach to consolidate ESG metrics into one central Fact_Company_Metrics table and link with other dimensions to the primary key of “Company.” Each dimensions table, starting from Dim_Company, provides contextual information for the company’s metadata. Dim_Industry discusses industry classifications for the company. Dim_Region stores geographical information and Dim_Time holds time-related data for temporal analysis.

As for the multiplicity, since one company can only have one company metadata, it holds a relationship of one-to-one. For the other components, each company can have multiple metrics over different periods, different metrics analyzed within industries, and metrics can aggregate across regions accordingly. Therefore, the rest holds a one-to-many relationship for the multiplicity.

5.2. Justification in how the data warehouse achieves key queries

With the uniqueness of star schema design, minimal joins can be used to effectively answer potential types of queries that are relevant to ESG. Below is a short list of potential queries that might arise from investors and can be effectively addressed through the constructed data warehouse.

1. Environmental Metrics:
   a. Average Carbon Emissions per year for each industry
   b. Total energy consumption by region
2. Social Metrics:
   a. Employee turnover rates per year by company
   b. Correlation between DEI scores and employee satisfaction
3. Governance Metrics:
   a. Relationship between governance rating and executive compensation
   b. Analysis of Shareholder Voting Participation Trends
4. Cross-Domain Insights:
   a. Trends in overall ESG rating over time
   b. Regional and industry-specific comparisons for ESG metrics.

5.3. Integration of real source into the DW

Issue Name	ESG Risk Exposure Score | Category	ESG Risk Management Score | Category	ESG Risk Rating Score | Category	Contribution to ESG Risk Rating
Corporate Governance	9.0 High	40.2 Average	**5.4 Medium**	23.1%
Carbon -Products and Services	6.4 Medium	17.0 Weak	**5.3 Medium**	22.8%
Human Capital	6.0 Medium	30.3 Average	**4.3 Medium**	18.4%
Product Governance	10.0 High	75.0 Strong	**2.5 Low**	10.7%
Business Ethics	6.0 Medium	63.7 Strong	**2.4 Low**	10.2%
Carbon -Own Operations	3.5 Low	42.5 Average	**2.0 Negligible**	8.5%
Human Rights -Supply Chain	2.0 Low	35.0 Average	**1.4 Negligible**	6.2%
Overall	**42.9 Medium**	**46.8 Average**	**23.3 Medium**	**100.0%**

Figure 5.2 - Vinfast Trading & Investment Pte Ltd ESG Statistics

The incorporation of VinFast Trading & Investment Pte Ltd. into the established data warehouse can be done through the process of mapping its measurable metrics and written content into the existing framework. Key metrics such as ESG Risk Rating (23.3 - Medium Risk), Environmental Exposure (e.g., Carbon Management Score of 17.0 - Weak), Social Factors (e.g., Human Capital Management at 30.3 - Average), and Governance Indicators (e.g., Business Ethics Management at 63.7 - Strong) can be directly loaded into the Fact_Company_Metrics table under the unique

Company_ID for VinFast. Dimensions like Dim_Company will store VinFast's metadata, such as its industry (Automobiles), region (Asia), and headquarters (Singapore), while Dim_Time captures the assessment year (2022). Unstructured data from the report will be further processed through NLP technique which will further enhance warehouses's ability to analyze textual, self-announced reports. Metrics that is related to environmental and social dimensions can easily be mapped into the schema for comprehensive analysis. Such demonstration displays the easy implementation of current available ESG statistics into the built DW from this project.

6. Uniquness and Donwfall feature of DW

There are a total of 5 unique features of the constructed data warehouse. The first unique aspect of the DW is that DW is constructed by placing the company as the center of the schema. Such design allows all metrics and dimensions to connect back to the primary entity, easing the process of corporate-level ESG analysis. This will enable investors to strictly focus on company-specific performance while supporting information can help broaden their analysis. Another uniqueness comes from integrated ESG metrics as metrics for environment, social, and governance have been integrated into one single fact table which allows holistic ESG evaluations. This will also hold the benefit of having an easier approach for cross-domain insights. The DW is incorporated with both structured (e.g. ESG scores, carbon emissions) and derived metrics (e.g. compliance default rates) which provides a wider range of analytical needs. The flexibility for both raw metric analysis and trend evaluation is a key unique feature. On top of this, the constructed design for the DW, which is the Kimball start schema, minimises join complexity, allowing it to have a highly optimized query performance. Large datasets can be easily accessed with fast query response times, making it practical for real-time dashboards. The last unique feature will be its scalability for future data. The design of the schema allows the introduction of new metrics dimensions easily. This means further updates on new data for the data warehouse can easily be employed without any computational barriers.

However, there are also some downfalls features of the constructed DW. The biggest downfall will be its limited flexibility for real-time updates. Since all metrics are consolidated into a single fact table, handling extremely high volumes of ESG data from past to recent, may cause performance bottlenecks in some situations, especially slower query time for large datasets. Another downfall will be the initial development complexity with the need to integrate diverse data sources and ensuring their quality requires significant time and effort before construing the data warehouse. This delays the implementation and has the potential to increase costs if not done effectively. Lastly, there can still be room for data not filtered out properly if such metrics have spanned multiple periods. This could potentially lead to data duplication and affect performance for large datasets.

6.1. Comparison with Competitors

The constructed ESG Data Warehouse offers an effective and creative approach to counter the limitations posed by the existing competitors, including S&P's ESG data management solution. One such distinctive quality is the confluence of structured and unstructured dimensions enabling the processing of quantitative (e.g., carbon consumption, standard of governance) and qualitative information from ESG documents or legal filings. The constructed warehouse for this project can recognise irregularities in governance structures, conduct reviews on sustainability reports, and perform basic analysis on ESG-related documents whether filled by the companies or external agencies, an aspect that S&P's warehouse can not achieve.

A further benefit is a company-oriented structure where the centre of this data warehouse's model is the company itself. This facilitates corporate analysis as companies' metrics and dimensions, time, industry, and even region are all easily connected to the company and so allow for a more detailed drill down. So companies both can invest in specific companies while looking at the general picture. This approach is also helpful as components of ESG ratings can be rated. This enables stakeholders to understand where the high or low scores are driven from. Such a unique structure will not be common in other solutions for ESG data warehouses.

The inclusion of real-time data management services in the warehouse also makes it distinct from traditional solutions which are primarily batch-oriented. The inclusion of Lambda architecture into the data warehouse allows the users to perform both historical processing of ESG data trends and real-time streaming of ESG data updates. Such functionality is essential while dealing with changing figures. Competitors, such as S&P, heavily rely on periodic ESG updates which does not help the users if they require immediate results. Therefore, investors always look for insights to balance risks and opportunities during both real-time and in different cycles, the constructed DW is much stronger compared to its competitors.

The warehouse also offers customizability and scalability features that allow the development of custom key point indicators as the needs of the stakeholders grow. This versatility empowers investors to create and study efficient metrics like 'carbon emissions for every dollar of revenue' or 'year-to-year DEI growth rates'. The design enables the use of various types of ESG sources including a combination of free data resources as well as paid ones like Sustainalytics and MSCI, making it inexpensive without sacrificing good quality of information. Therefore, this characteristic of the warehouse system makes it a relevant future-proof solution relative to the ones existing currently.

In this regard, the uniqueness of the DW for this project cannot be replicated by solutions from S&P and others. Its non-linear analytic architecture, real-time processing capabilities, and many other features excel among its competitors. These advantages imply that the developed warehouse is not only the suitable system available to address current ESG-related issues but also an expandable geographic information system that will be useful for future investments.

7. Conclusion

The ESG Data Warehouse as proposed in this project provides a solid foundation for buy-side investors to perform assessments of companies based on Environmental, Social and Governance factors. Despite some downfalls, the well-addressed ESG Data Warehouse delivers real value.

Overall, this project displays an exemplified example of how data engineering can solve potential real-world problems. It provides details about effective methods of ESG data management with practical solutions for investors navigating the growing importance of sustainability for investment.

8. Sources

ESG (environmental, Social, & Governance). Corporate Finance Institute. (2023, October 26). https://corporatefinanceinstitute.com/resources/esg/esg-environmental-social-governance/

M&T Investments Pte. ltd.. (n.d.). https://mstar-sustops-cdn-mainwebsite-s3.s3.amazonaws.com/docs/default-source/spos/m-t-investments-pte-ltd-annual-review-(2024).pdf?sfvrsn=fb35b9c5_1

Morgan, L. (2022, January 26). *NLP and AI boost the Automated Data Warehouse: TechTarget*. Search Data Management. https://www.techtarget.com/searchdatamanagement/feature/NLP-and-AI-boost-the-automated-data-warehouse