École Nationale Supérieure
d'Informatique et d'Analyse des Systèmes
- Rabat

Projet BI

BBS Statistical DW : Industry Business Wing

Élèves :
Ayman Fenkouch
Yahya Moudrik
Hiba Asghar
Youssef Rizki

Encadrant :
Pr. Benhiba

Année Universitaire 2025-2026

Abstract

This report presents the design and implementation of the Industry & Business module within the Bangladesh Bureau of Statistics (BBS) national statistical data warehouse project. In collaboration with the Korea International Cooperation Agency (KOICA), this initiative addresses the critical challenge of fragmented statistical data across multiple governmental institutions by developing a centralized Business Intelligence platform.
The project encompasses three interconnected analytical domains : Industry, Internal Business, and External Business. Each domain was implemented as a dedicated datamart following dimensional modeling principles, enabling comprehensive analysis of Bangladesh's industrial performance, domestic market dynamics, and international trade flows. The technical implementation leveraged Microsoft BI technologies, including SQL Server for data warehousing, SQL Server Integration Services (SSIS) for ETL processes, and Power BI for interactive dashboards.
Key achievements include the successful integration of heterogeneous data sources, the design of multidimensional schemas supporting complex analytical queries, and the development of user-friendly dashboards delivering actionable insights for strategic decision-making. This work demonstrates the practical application of data warehousing methodologies in supporting national statistical infrastructure and evidence-based policymaking.

Introduction

Modern governance and economic development increasingly depend on timely, accurate, and accessible statistical information. In Bangladesh, the challenge of managing national statistics has been compounded by the dispersion of data across multiple ministries, public institutions, and directorates. These data sources exhibit heterogeneous formats, inconsistent indicator definitions, limited historical depth, and weak interoperability, significantly hindering comprehensive cross-domain analysis and strategic decision support.

Recognizing these challenges, the *Bangladesh Bureau of Statistics* (BBS), in partnership with the *Korea International Cooperation Agency* (KOICA), has undertaken an ambitious national modernization program aimed at transforming the country's statistical information system. At the heart of this initiative lies the **Statistical Data Warehouse Project**, designed to establish a centralized Business Intelligence platform capable of collecting, integrating, storing, and analytically exploiting national statistical data.

The overall system architecture is organized into thematic BI modules, each addressing a major statistical domain. Our team has been entrusted with the development of the **Industry & Business module**, which plays a critical role in measuring economic performance, industrial productivity, internal market dynamics, and external trade flows.

This module consolidates three interconnected analytical subdomains :

1. **Industry Submodule** : Focuses on measuring the structure, performance, and economic contribution of industrial activities, providing insights into productivity, employment, investment, and value creation across the industrial sector.
2. **Internal Business Submodule** : Analyzes domestic market dynamics, enterprise performance, and commercial activity, supporting monitoring of business ecosystem health, entrepreneurial activity, and internal economic vitality.
3. **External Business Submodule** : Examines international trade flows, capturing imports, exports, trade balance, and external competitiveness indicators essential for understanding Bangladesh's position in global markets.

The Industry & Business module serves multiple strategic objectives. It supports policymakers in making evidence-based decisions regarding industrial development, trade policy, and economic planning. It facilitates the monitoring of national development plans and ensures compliance with international reporting standards, including the Sustainable Development Goals (SDGs). Furthermore, it improves public access to official statistics, promoting transparency and informed civic engagement.

Our implementation follows established Business Intelligence best practices and data warehousing principles. The architecture is built upon dimensional modeling using star schemas, ensuring optimal query performance and analytical flexibility. Extract, Transform, Load (ETL) processes were designed to ensure data quality, consistency, and traceability throughout the integration pipeline. Interactive dashboards and analytical reports were developed to translate complex data into actionable insights accessible to diverse stakeholder groups.

This report documents the complete lifecycle of the Industry & Business module development, from initial analysis through final implementation. Chapter **??** presents a

comprehensive analysis of the project context, requirements, and domain specifications. Chapter **??** details the conceptual and logical design of each datamart, including dimensional modeling and dashboard prototypes. Chapter **??** describes the technical realization, covering data generation, ETL implementation, data warehouse construction, and dashboard development. The report concludes with reflections on achievements, challenges encountered, and recommendations for future enhancements.

Through this work, we demonstrate how modern Business Intelligence technologies and methodologies can be effectively applied to strengthen national statistical infrastructure, ultimately contributing to more informed governance and sustainable economic development in Bangladesh..

Table des matières

Abstract **1**

Introduction **2**

1 Analyse **7**
1.1 Context of the Project . . . 7
1.2 Project Objectives . . . 8
1.2.1 Functional Requirements . . . 8
1.2.2 Technical Requirements . . . 8
1.3 Industry & Business Module Description . . . 9
1.3.1 Industry Submodule . . . 9
1.3.2 Internal Business Submodule . . . 10
1.3.3 External Business Submodule . . . 11
1.4 Multidimensional Matrix . . . 12
1.4.1 Industry SubModule . . . 12
1.4.2 Internal Trade SubModule . . . 12
1.4.3 External Trade SubModule . . . 12

2 Conception **13**
2.1 Industry . . . 13
2.1.1 Datamart Objectives . . . 13
2.1.2 Key Performance Indicators (KPIs) . . . 13
2.1.3 Datamart Dimensions . . . 14
2.1.4 Datamart Modeling . . . 14
2.1.5 Report and Dashboard Design . . . 14
2.2 Internal Business . . . 16
2.2.1 Datamart Objectives . . . 16
2.2.2 Key Performance Indicators (KPIs) . . . 16
2.2.3 Datamart Dimensions . . . 16
2.2.4 Datamart Modeling . . . 17
2.2.5 Report and Dashboard Design . . . 17
2.3 External Business . . . 19
2.3.1 Datamart Objectives . . . 19
2.3.2 Key Performance Indicators (KPIs) . . . 19
2.3.3 Datamart Dimensions . . . 19
2.3.4 Datamart Modeling . . . 19
2.3.5 Report and Dashboard Design . . . 20

3 Realisation **21**

3.1 Industry Sector 21
3.1.1 Data Generation 21
3.1.2 Staging Tables and Data Mart 21
3.1.3 ETL Process Using SSIS 22
3.1.4 Data Restitution and dashboard Building : 24
3.2 Internal Business 26
3.2.1 Technology choices and working environment 26
3.2.2 Synthetic data generation 26
3.2.3 Adopted ETL architecture 27
3.2.4 Loading the staging area 27
3.2.5 Design and implementation of the data warehouse 28
3.2.6 Fact table loading 29
3.2.7 Validation and execution of processes 30
3.2.8 Power BI Dashboards 31
3.3 External Business 33
3.3.1 DataMart Physical Design 33
3.3.2 ETL Conception 34
3.3.3 ETL Development Using SSIS 35
3.3.4 User Application : Power BI Dashboards 36

Conclusion **38**

Table des figures

1.1 Industry MultiDimensional Matrix 12
1.2 Internal Business MultiDimensional Matrix 12
1.3 External Business MultiDimensional Matrix 12

2.1 Industry Datamart Schema 14
2.2 Industry Dashboard Mockup 15
2.3 Industry Dashboard Mockup 15
2.4 Internal Business Datamart Schema 17
2.5 Internal Business Dashboard Mockup 17
2.6 Internal Business Dashboard Mockup 18
2.7 External Business Datamart Schema 19
2.8 External Business Dashboard Mockup 20

3.1 Assets Overview Dashboard 22
3.2 Load data from raw source to the staging area 22
3.3 Load data from staging area to the dimension tables 23
3.4 Load data from staging area to the fact tables 23
3.5 Main package 23
3.6 Industry Overview Dashboard 24
3.7 Industry Workforce Overview 25
3.8 Assets Overview Dashboard 25
3.9 Example of an SSIS data flow for loading data into the staging area 28
3.10 Star schema of the Internal Business data warehouse 29
3.11 SSIS data flow for loading the Internal Business fact table 29
3.12 Detailed SSIS data flow for fact table processing 30
3.13 Successful execution of the ETL process and fact table loading 31
3.14 Internal Business and Industry Dashboard 1 32
3.15 Internal Business and Industry Dashboard 2 32
3.16 Physical DataMart in SQL Server 34
3.17 External Business ETL in SSIS 36
3.18 External Business Dashboard in Power BI 37

Chapitre 1

Analyse

1.1 Context of the Project

The Bangladesh Bureau of Statistics (BBS), in collaboration with the Korea International Cooperation Agency (KOICA), has initiated a large-scale national modernization program aimed at strengthening the statistical information system of Bangladesh.

Currently, statistical data related to industry, trade and business activities are distributed across several ministries, public institutions and directorates. These data sources are characterized by heterogeneous formats, inconsistent definitions of indicators, limited historical depth and weak interoperability, which significantly limits cross-domain analysis and decision support.

To address these challenges, the **Statistical Data Warehouse Project** aims to design and implement a centralized Business Intelligence (BI) platform that enables the collection, integration, storage and analytical exploitation of national statistical data. The project relies on a modern BI architecture based on Data Warehousing principles, ETL processes, OLAP cubes and interactive dashboards.

This platform is intended to support strategic decision-making for policymakers, facilitate monitoring of national development plans, ensure compliance with international reporting standards (including SDGs), and improve public access to official statistics.

Within this framework, the overall system is structured into thematic BI modules, each corresponding to a major statistical domain. Our group is responsible for the **Industry & Business module**, which plays a critical role in measuring economic performance, industrial productivity, internal market dynamics and external trade flows.

1.2 Project Objectives

The Industry & Business BI module must comply with the following functional and technical requirements :

1.2.1 Functional Requirements

— Integrate data from multiple heterogeneous sources related to industry, internal trade and external trade.
— Harmonize statistical definitions, nomenclatures and classifications (ISIC, HS codes, geographic codes, time granularity).
— Provide a multidimensional analysis framework enabling slicing, dicing, drill-down and roll-up operations.
— Deliver reliable and validated indicators for national monitoring and reporting.
— Support historical analysis and trend monitoring.
— Enable automated generation of statistical reports and dashboards.

1.2.2 Technical Requirements

— Design a dedicated **Datamart** for the Industry & Business domain.
— Implement robust **ETL processes** ensuring data quality, consistency and traceability.
— Build **OLAP cubes** to support complex analytical queries.
— Develop interactive dashboards and reporting views using Microsoft BI tools.
— Ensure scalability to support future data sources and indicators.

1.3 Industry & Business Module Description

The **Industry & Business** module consolidates all industrial, trade and enterprise-related indicators needed for national monitoring. It integrates the following subdomains :

1.3.1 Industry Submodule

The Industry submodule focuses on measuring the structure, performance and economic contribution of industrial activities at national and regional levels. It provides key indicators required for monitoring productivity, employment, investment and value creation in the industrial sector.

Key Indicators / KPIs

- **Number of industrial establishments** : Measures the total number of active industrial units, providing insight into the size, density and structural evolution of the industrial sector.
- **Total employment** : Represents the total number of workers employed in industrial establishments, reflecting the sector's contribution to job creation.
- **Total wages and salaries** : Captures the total compensation paid to employees, allowing analysis of labor costs and income distribution within the industrial sector.
- **Industrial fixed assets value** : Measures the value of long-term tangible assets (machinery, buildings, equipment), indicating the level of industrial investment and production capacity.
- **Capital stock** : Represents the accumulated value of productive assets over time, used to assess industrial sustainability and long-term growth potential.
- **Raw material consumption** : Quantifies the volume or value of raw materials used in production, enabling analysis of production efficiency and input dependency.
- **Energy consumption** : Measures energy usage by industrial activities, supporting productivity analysis as well as environmental and energy-efficiency assessments.
- **Intermediate consumption** : Represents the value of goods and services consumed during the production process, excluding fixed assets, and is essential for value-added calculations.
- **Gross Output** : Measures the total value of goods produced by the industrial sector before deducting intermediate consumption.
- **Gross Value Added (GVA)** : Represents the net contribution of industry to the economy, calculated as Gross Output minus Intermediate Consumption.
- **Industrial Tax** : Measures taxes paid by industrial activities, providing insight into fiscal contribution and public revenue generated by the sector.

Analysis Axes

- **Time** : Enables temporal analysis of industrial performance across years, quarters or months to identify trends and cycles.
- **Industry classification (ISIC)** : Allows comparison and aggregation of indicators by standardized industrial activity categories.

— **Establishment size** : Segments industrial units by size (small, medium, large) to analyze structural differences and productivity levels.
— **Ownership type** : Distinguishes between public, private and foreign-owned establishments to assess ownership impact on performance and investment.
— **Geographic location** : Enables spatial analysis at division or district level to identify regional disparities and industrial concentration.

1.3.2 Internal Business Submodule

The Internal Business submodule analyzes domestic market dynamics, price behavior and commercial activity, supporting the monitoring of inflation, supply conditions and market performance.

Key Indicators / KPIs

— **Number of active enterprises** : Measures the total number of enterprises currently operating in the domestic economy, serving as a core indicator of economic activity.
— **Enterprise density** : Represents the number of enterprises relative to population or geographic area, enabling assessment of business concentration and regional economic vitality.
— **Enterprise size distribution** : Shows the breakdown of enterprises by size (micro, small, medium, large), providing insight into the structure of the productive fabric.
— **Domestic turnover** : Measures total revenue generated by enterprises in the domestic market, reflecting internal market performance.
— **Average turnover per enterprise** : Indicates the mean revenue per enterprise, allowing productivity and scale comparisons across sectors and regions.
— **Domestic trade volume** : Captures the total volume of goods and services exchanged within the domestic market, reflecting internal commercial flows.
— **Business creation count** : Measures the number of newly created enterprises over a given period, indicating entrepreneurial dynamism.
— **Business closure count** : Tracks the number of enterprises that ceased activity, helping assess market exits and economic stress.
— **Net business growth** : Represents the balance between enterprise creations and closures, providing a synthetic indicator of business ecosystem expansion or contraction.

Analysis Axes

— **Time** : Enables analysis of enterprise dynamics, turnover evolution, and structural changes over different periods.
— **Region** : Supports territorial analysis of business activity, highlighting regional disparities and development patterns.
— **Business sector** : Allows sectoral analysis to compare performance, growth, and resilience across different branches of economic activity.
— **Enterprise size** : Distinguishes enterprises by size class, enabling assessment of structural composition and contribution by firm scale.

— **Enterprise status** : Differentiates enterprises by operational status (active, newly created, closed), supporting lifecycle and survival analysis.
— **Market** : Enables segmentation by market orientation (domestic market segments), allowing deeper analysis of internal commercial activity.

1.3.3 External Business Submodule

The External Business submodule focuses on international trade flows, external competitiveness and economic openness.

Key Indicators / KPIs

— **Total imports value (USD)** : Measures the total value of goods imported, reflecting dependency on foreign markets.
— **Total exports value (USD)** : Measures the total value of exported goods, indicating international market performance.
— **Import growth rate** : Captures the rate of change in imports over time, highlighting trade expansion or contraction.
— **Export growth rate** : Measures export performance dynamics and competitiveness.
— **Trade balance** : Represents the difference between exports and imports, indicating surplus or deficit.
— **Import share of GDP** : Measures the relative importance of imports in the national economy.
— **Export share of GDP** : Measures the contribution of exports to economic output.

Analysis Axes

— **Time** : Enables temporal analysis of trade trends and seasonality.
— **Product classification (HS code)** : Allows standardized analysis by product type in international trade.
— **Trade flow** : Distinguishes imports from exports for directional analysis.
— **Partner country** : Enables analysis of trade relationships and geographic diversification.
— **Product type** : Groups products into consumer goods, raw materials and capital goods.
— **Account type** : Differentiates trade by institutional sector (private, government, semi-government).
— **Transport mode** : Allows analysis of logistics patterns and transport infrastructure usage.

1.4 Multidimensional Matrix

1.4.1 Industry SubModule

		Time	Establishment	Employment	AssetType
Industry	IndustrialCost	x	x		
	NonIndustrialCost	x	x		
	IntermediateConsumption	x	x		
	GrossOutput	x	x		
	IndustrialTax	x	x		
	RawMaterialCost	x	x		
	EnergyCost	x	x		
	TotalPersonEngaged	x	x	x	
	SalaryandWages	x	x	x	
	CashBenefits	x	x	x	
	NonCashBenefits	x	x	x	
	SocialSecurityCost	x	x	x	
	OpeningValue	x	x		x
	Depreciation	x	x		x
	CapitalExpenditure	x	x		x
	NetFixedAssets	x	x		x

FIGURE 1.1 – Industry MultiDimensional Matrix

1.4.2 Internal Trade SubModule

		Time	Region	Business Sector	Enterprise Size	Enterprise Status	Market
Internal Business	No Active Enterprises	x	x	x	x	x	x
	Enterpresie Density	x	x	x	x	x	x
	Enterprise Size	x	x	x	x	x	x
	Domestic Turnover	x	x	x	x	x	x
	Avg Turnover	x	x	x	x	x	x
	Domestic Trade Volume	x	x	x	x	x	x
	Business Creation Count	x	x	x	x	x	x
	Business Closure Count	x	x	x	x	x	x
	Net Business Growth	x	x	x	x	x	x

FIGURE 1.2 – Internal Business MultiDimensional Matrix

1.4.3 External Trade SubModule

		Time	Region	Composition	Account	Routes
External Business	Imports	x	x	x	x	x
	Exports	x	x	x	x	x

FIGURE 1.3 – External Business MultiDimensional Matrix

Chapitre 2

Conception

2.1 Industry

2.1.1 Datamart Objectives

The Industry Datamart aims to analyze industrial performance in Bangladesh through key statistical indicators related to production, value added, employment, and sectoral growth.

2.1.2 Key Performance Indicators (KPIs)

— **IndustrialCost** : Represents the total costs directly related to industrial production activities, including manufacturing operations and production processes.
— **NonIndustrialCost** : Captures costs not directly linked to production, such as administrative, marketing or support services expenses.
— **IntermediateConsumption** : Measures the value of goods and services consumed during production, excluding fixed assets, and is a core component in value-added calculation.
— **GrossOutput** : Represents the total value of goods produced before deducting intermediate consumption.
— **IndustrialTax** : Measures taxes paid by industrial establishments, reflecting their fiscal contribution to public revenue.
— **RawMaterialCost** : Represents the total cost of raw materials used in the production process, enabling analysis of input dependency and production efficiency.
— **EnergyCost** : Measures expenditure on energy consumption, supporting analysis of energy intensity and cost structure.
— **TotalPersonEngaged** : Represents the total number of persons engaged in industrial activity, including employees and working owners.
— **SalaryandWages** : Captures total monetary compensation paid to employees.
— **CashBenefits** : Measures additional cash-based employee benefits beyond salaries and wages.
— **NonCashBenefits** : Represents the value of in-kind benefits provided to employees, such as housing or transportation.
— **SocialSecurityCost** : Captures employer contributions to social security and insurance schemes.
— **OpeningValue** : Represents the opening balance value of fixed assets at the beginning of the accounting period.
— **Depreciation** : Measures the reduction in value of fixed assets due to wear, obsolescence or aging.
— **CapitalExpenditure** : Represents investments in new or improved fixed assets during the period.

— **NetFixedAssets** : Measures the net value of fixed assets after depreciation.

2.1.3 Datamart Dimensions

— **Time Dimension** : Enables temporal analysis of industrial indicators by year, quarter or month, supporting trend and growth analysis.
— **Industrial Class Dimension** : Represents industrial activity classification based on standardized codes (e.g. ISIC), enabling sector-level comparison and aggregation.
— **Establishment Class Dimension** : Segments establishments by size or category, supporting structural analysis between small, medium and large units.
— **Employment Type Dimension** : Distinguishes types of employment (permanent, temporary, skilled, unskilled), allowing workforce structure analysis.
— **Assets Type Dimension** : Classifies assets by type (machinery, buildings, vehicles), supporting detailed capital structure analysis.

2.1.4 Datamart Modeling

The following figure presents the logical schema of the Industry Datamart.

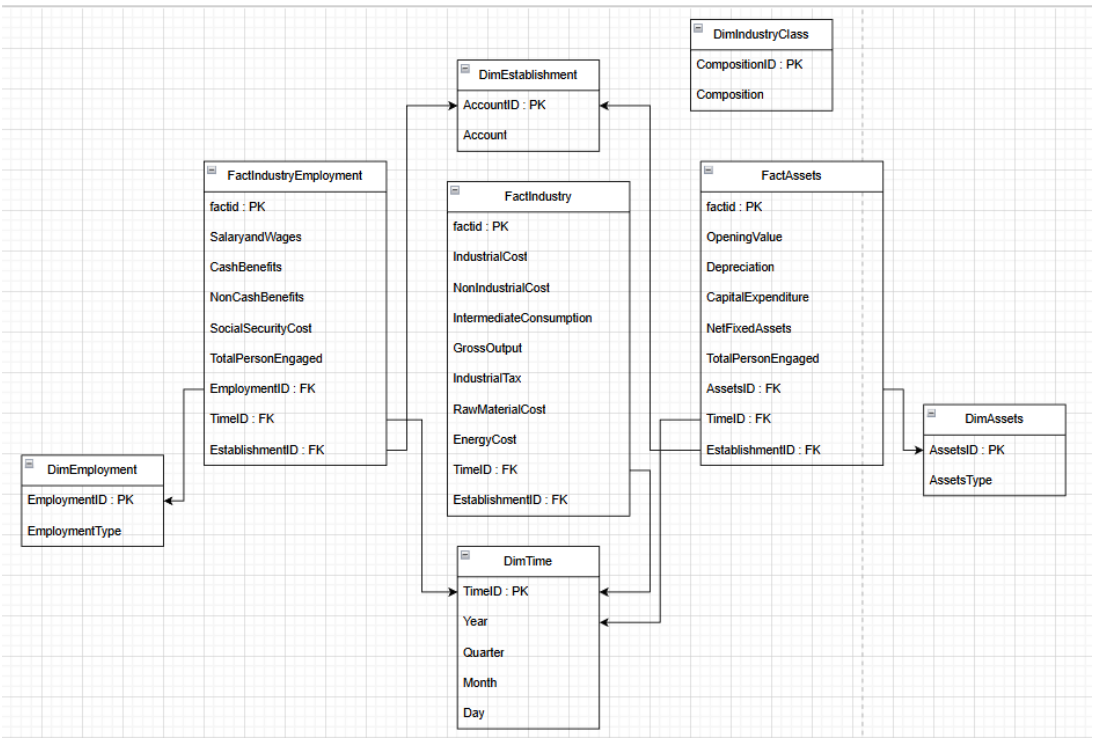

FIGURE 2.1 – Industry Datamart Schema

2.1.5 Report and Dashboard Design

The reports and dashboards allow monitoring industrial performance across time, Establishments, and classes.

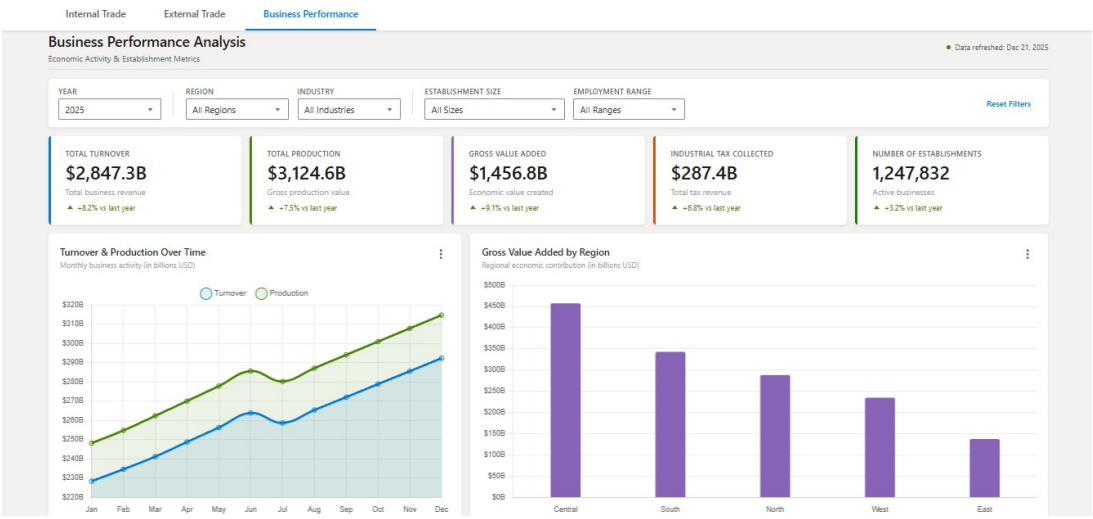

FIGURE 2.2 – Industry Dashboard Mockup

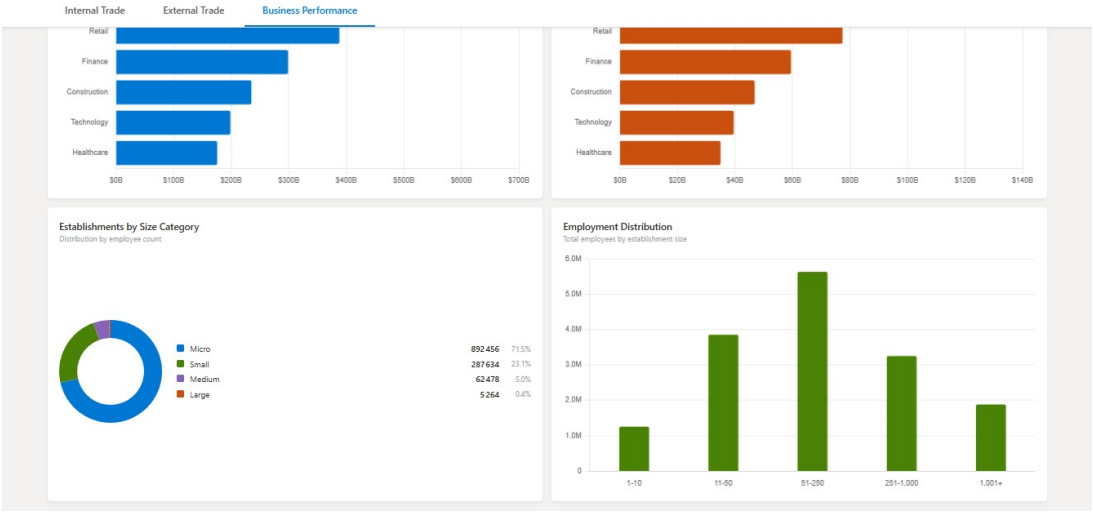

FIGURE 2.3 – Industry Dashboard Mockup

2.2 Internal Business

2.2.1 Datamart Objectives

The Internal Business Datamart focuses on the analysis of domestic economic activities, particularly the performance of local enterprises and their contribution to the national economy.

2.2.2 Key Performance Indicators (KPIs)

- **ActiveEnterprises** : Represents the total number of enterprises operating in the domestic economy, serving as a core indicator of internal business activity.
- **EnterpriseDensity** : Measures the concentration of enterprises relative to population or geographic area, enabling assessment of business distribution and regional vitality.
- **DomesticTurnover** : Captures the total revenue generated by enterprises in the domestic market, reflecting internal market performance.
- **AverageTurnover** : Represents the mean turnover per enterprise, allowing productivity and scale comparisons across sectors and regions.
- **DomesticTradeVolume** : Measures the total volume of goods and services exchanged within the domestic market, reflecting internal commercial flows.
- **BusinessCreations** : Counts newly created enterprises during a given period, indicating entrepreneurial dynamism.
- **BusinessClosures** : Counts enterprises that ceased activity, providing insight into market exits and economic pressures.
- **NetBusinessGrowth** : Represents the balance between business creations and closures, summarizing overall business ecosystem expansion or contraction.

2.2.3 Datamart Dimensions

- **Time Dimension** : Enables temporal analysis of enterprise dynamics, turnover evolution, and structural changes.
- **Region Dimension** : Supports territorial analysis of business activity and regional disparities.
- **Business Sector Dimension** : Allows sector-based analysis of enterprise distribution, performance, and growth patterns.
- **Enterprise Size Dimension** : Classifies enterprises by size (micro, small, medium, large), enabling structural and contribution analysis by firm scale.
- **Enterprise Status Dimension** : Differentiates enterprises by lifecycle status (active, created, closed), supporting business demography analysis.
- **Market Dimension** : Enables segmentation by domestic market category, allowing deeper analysis of internal commercial activity.

2.2.4 Datamart Modeling

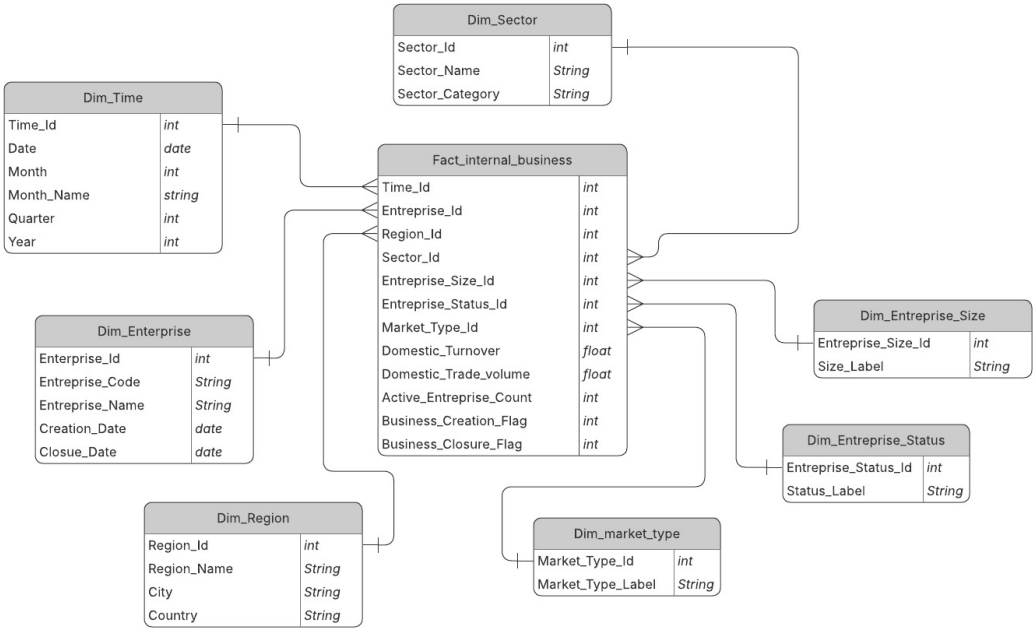

FIGURE 2.4 – Internal Business Datamart Schema

2.2.5 Report and Dashboard Design

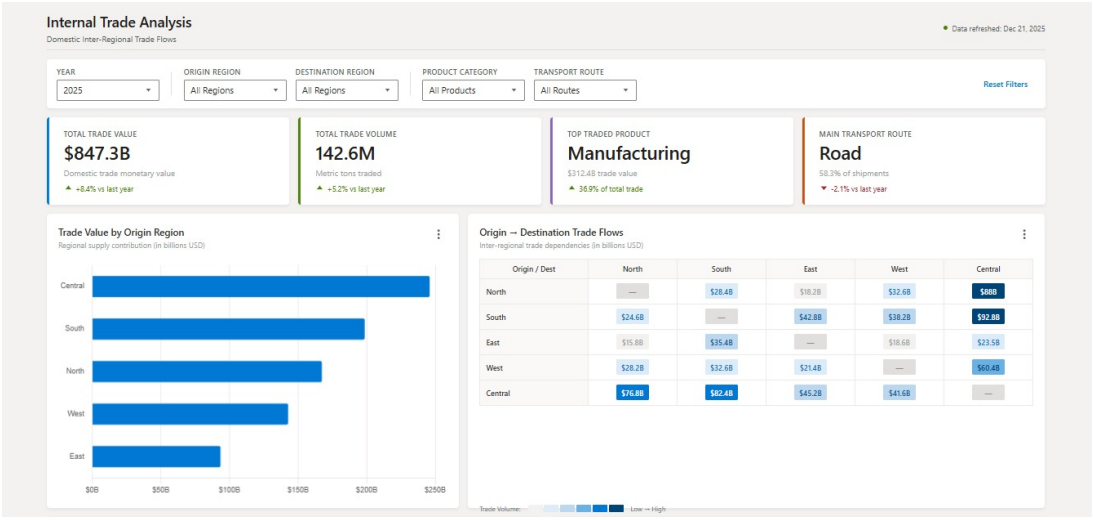

FIGURE 2.5 – Internal Business Dashboard Mockup

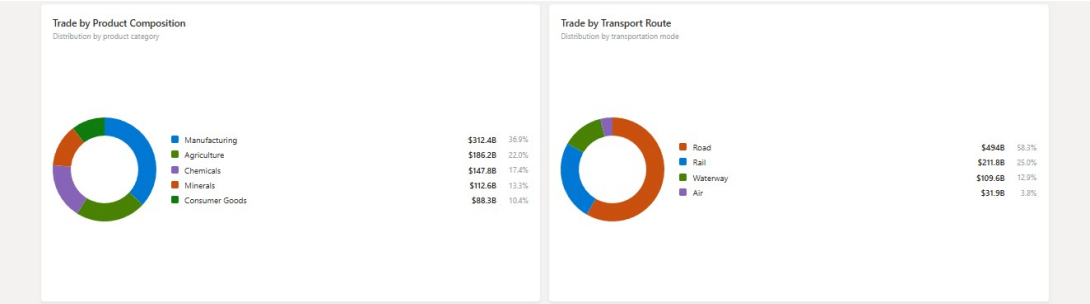

FIGURE 2.6 – Internal Business Dashboard Mockup

2.3 External Business

2.3.1 Datamart Objectives

The External Business Datamart is dedicated to the analysis of international trade activities, including imports, exports, and trade balance indicators.

2.3.2 Key Performance Indicators (KPIs)

— **Export Volume** : Measures the quantity or value of exported goods, indicating international competitiveness and external demand.
— **Import Volume** : Represents the quantity or value of imported goods, reflecting dependency on foreign markets.

2.3.3 Datamart Dimensions

— **Time Dimension** : Enables temporal monitoring of trade flows and seasonality.
— **Composition Dimension** : Classifies traded goods by product composition or category.
— **Account Dimension** : Differentiates trade flows by account type (private, government, semi-government).
— **Route Dimension** : Represents the transport route used for trade (sea, land, air), enabling logistics and infrastructure analysis.

2.3.4 Datamart Modeling

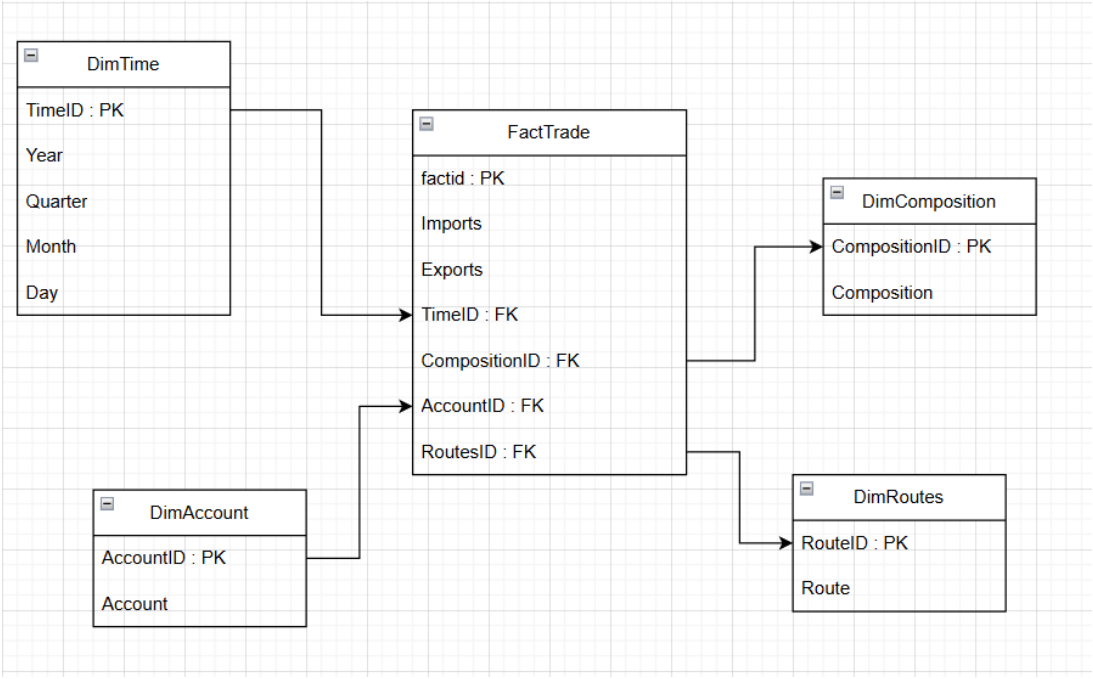

FIGURE 2.7 – External Business Datamart Schema

2.3.5 Report and Dashboard Design

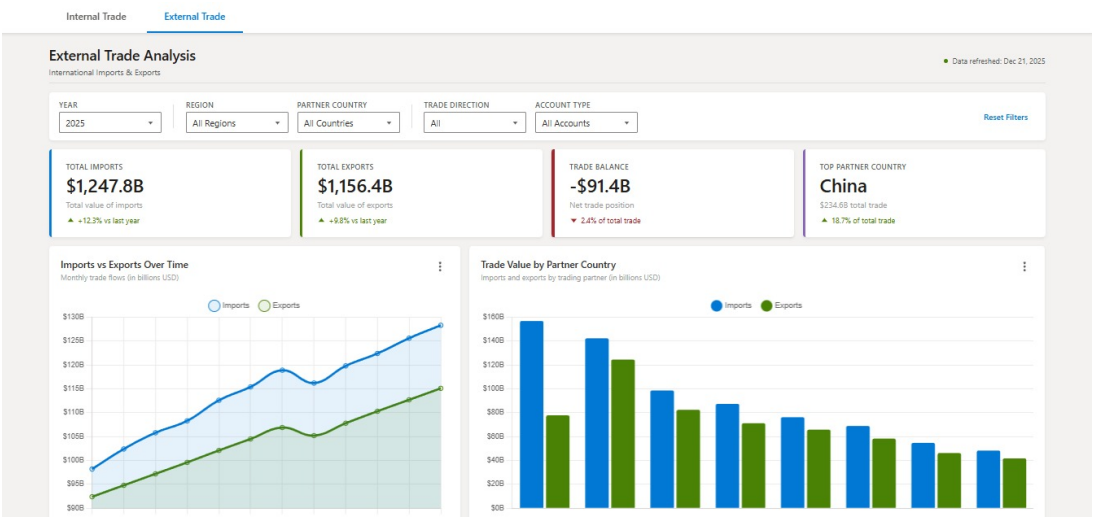

FIGURE 2.8 – External Business Dashboard Mockup

Chapitre 3

Realisation

3.1 Industry Sector

This section provides a comprehensive description of the technical implementation of the decision-support system developed for the **Industry Sector**. It presents the complete realization process, beginning with data acquisition and preprocessing, continuing through the ETL design and implementation, and culminating in the final population of the sector-specific data warehouse. Particular emphasis is placed on the technical choices, the adopted ETL architecture, and the transformations executed using SQL Server Integration Services (SSIS).

3.1.1 Data Generation

The data extraction process begins with the generation of structured datasets from the SMI document, which produces three CSV files containing employment, assets, and industrial data. These files serve as the raw input for the BI pipeline. Using SSIS, a dedicated extraction package was implemented to read the CSV sources, validate their structure, and load the contents into staging tables. This step ensures that the data is centralized, cleansed of basic inconsistencies, and made available in a controlled environment before any transformations are applied. By isolating the extraction phase, the solution guarantees reliability, reproducibility, and a clear separation between raw source data and the curated data mart.

3.1.2 Staging Tables and Data Mart

A staging area was implemented in SQL Server Management Studio (SSMS) with three staging tables to temporarily hold raw data. The data mart was designed using a star schema, consisting of fact tables (`Employments_Fact`, `Assets_Fact`, `Industrial_Fact`) and dimension tables (`Time_Dim`, `Establishment_Dim`, `IndustryClassification`, `Employment_Dim`).

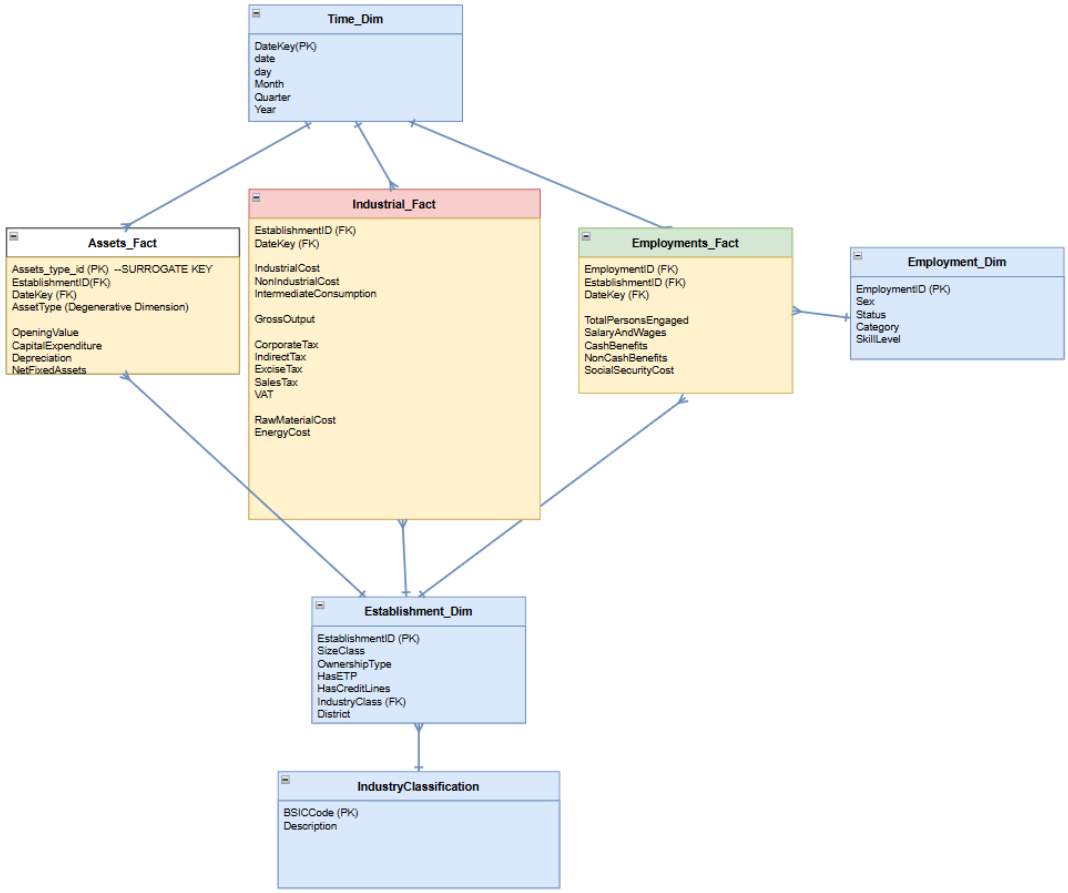

FIGURE 3.1 – Assets Overview Dashboard

3.1.3 ETL Process Using SSIS

The ETL pipeline was implemented with four SSIS packages :

— One package for extracting data from CSV files and loading it into the staging area.

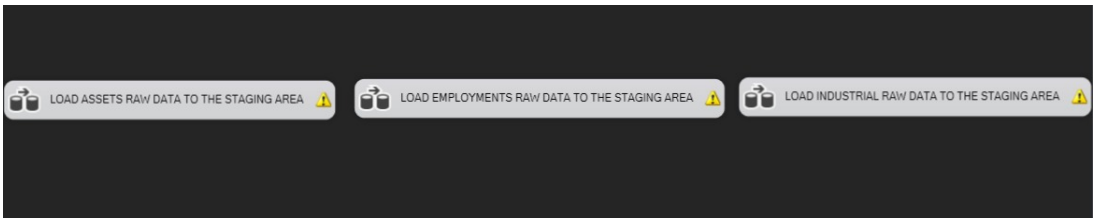

FIGURE 3.2 – Load data from raw source to the staging area

— One package for transforming data from staging and loading it into the dimension tables.

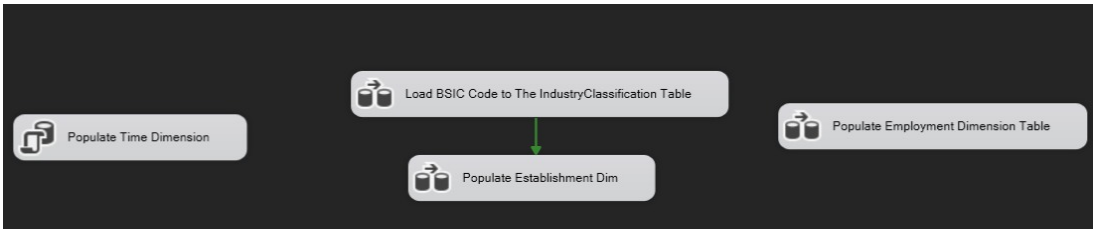

FIGURE 3.3 – Load data from staging area to the dimension tables

— One package for transforming data from staging and loading it into the fact tables.

FIGURE 3.4 – Load data from staging area to the fact tables

— One master package that executes the three previous packages sequentially using the Execute Package Task.

FIGURE 3.5 – Main package

3.1.4 Data Restitution and dashboard Building :

Once the data mart was populated, dashboards were created in Power BI to visualize industrial metrics. These include the :

— **Industry Sector Overview Dashboard**
This dashboard provides a high-level view of industrial performance. It highlights key metrics such as Gross Output, Total Taxes, and Total Energy Cost. A regional map shows the geographic distribution of outputs, while a bar chart ranks the top contributing sectors, with garment manufacturing as the leading industry. A pie chart further breaks down gross output by establishment size, showing that large enterprises dominate production.

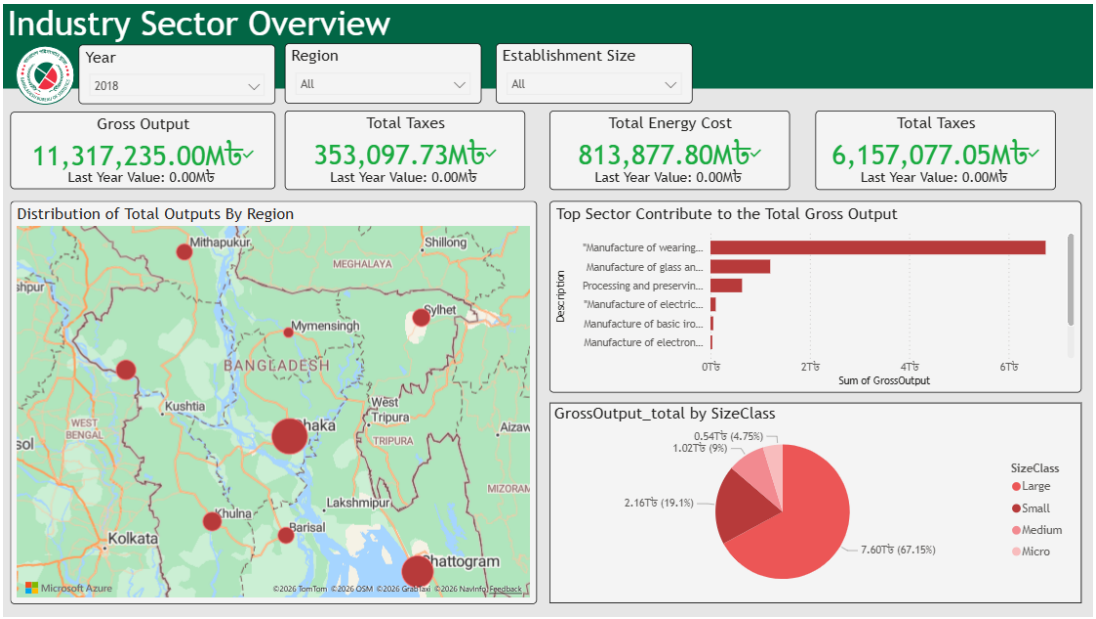

FIGURE 3.6 – Industry Overview Dashboard

— **Employment Distribution Dashboard**
This dashboard focuses on workforce structure and costs. It reports the total number of persons engaged, gender ratios, and average employment costs. Bar charts illustrate workforce distribution by sex, skill level, and employment category, revealing that most workers are skilled and concentrated in production roles. It also provides insights into administrative, clerical, and temporary employment categories.

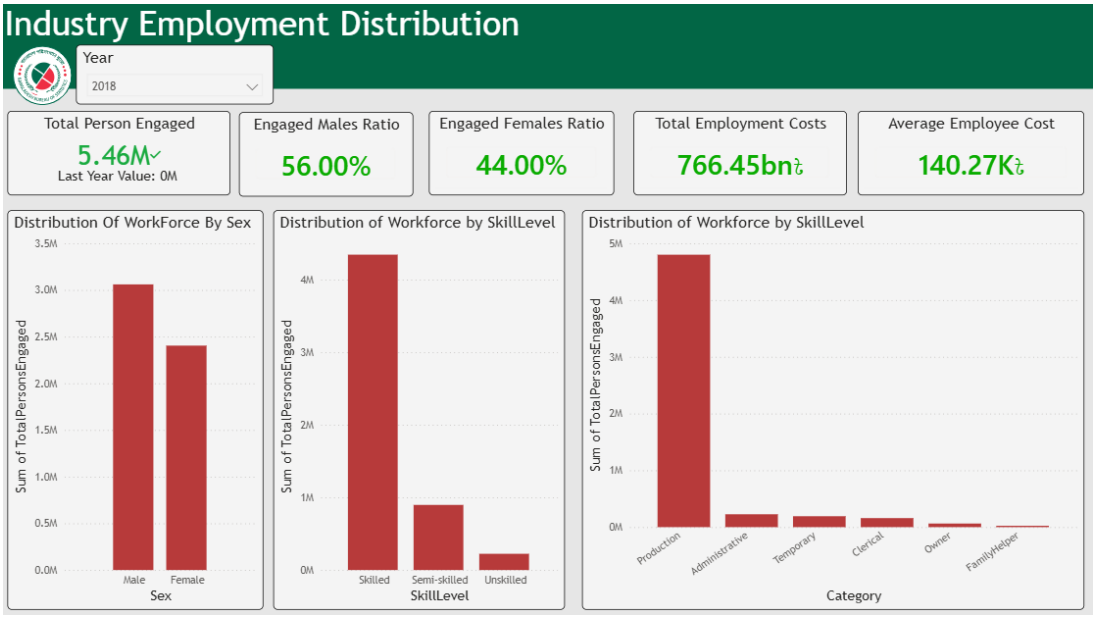

FIGURE 3.7 – Industry Workforce Overview

Assets Overview Dashboard

This dashboard presents financial insights into asset management. It displays total capital expenditure, net fixed assets, depreciation percentage, and the growth rate from opening to closing values. A bar chart shows capital expenditure by establishment size, with large enterprises investing the most. Another chart compares capital expenditure and depreciation across asset types, highlighting machinery as the dominant category.

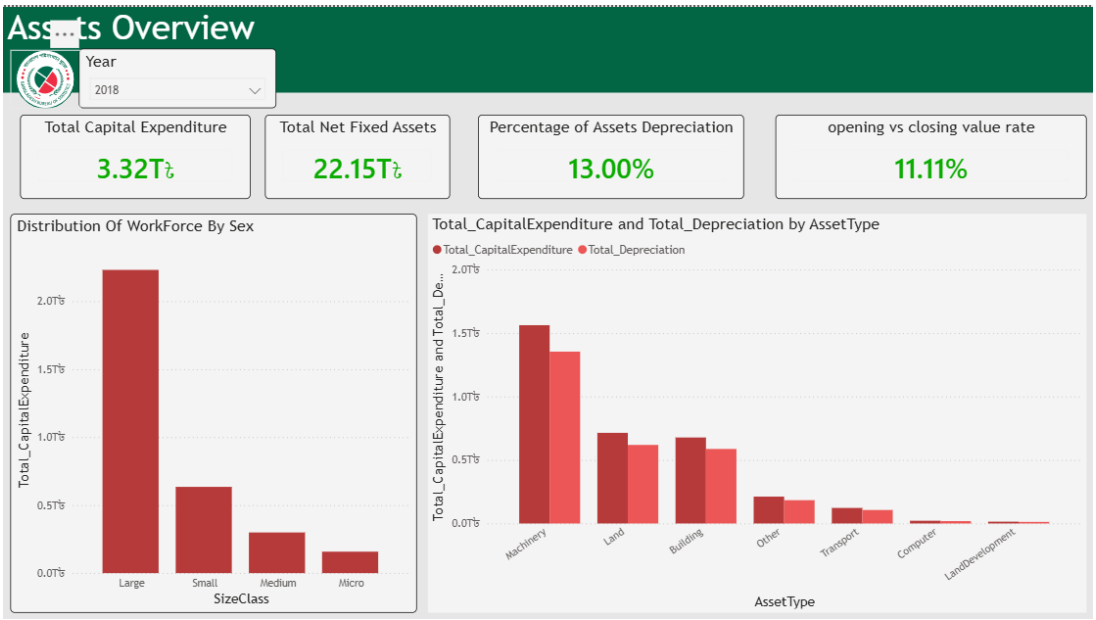

FIGURE 3.8 – Assets Overview Dashboard

.

3.2 Internal Business

This section provides a comprehensive description of the technical implementation of the decision-support system developed for the *Internal Business* sub-domain. It presents the complete realization process, starting from data generation and preparation, through the ETL design and implementation, and ending with the final population of the data warehouse. Particular emphasis is placed on the technical choices, the adopted ETL architecture, and the transformations performed using SQL Server Integration Services (SSIS).

The objective of this implementation is to transform raw operational data into structured, reliable, and analyzable information that supports decision-making related to internal economic activity in Bangladesh.

3.2.1 Technology choices and working environment

The implementation of the project is based on a Microsoft-centric environment, which is widely adopted in enterprise decision-support and business intelligence projects. This technological stack was selected for its robustness, integration capabilities, and relevance in real-world professional contexts.

The technologies used are :

— **Microsoft SQL Server**, used to manage relational databases for both the staging area and the data warehouse.

— **SQL Server Integration Services (SSIS)**, employed to design, implement, and execute ETL (Extract, Transform, Load) processes.

— **Visual Studio**, serving as the integrated development environment for building and managing SSIS packages.

— **CSV files**, chosen as the format for source data exchange due to their simplicity, portability, and widespread use.

This technological choice enables the construction of a complete and realistic decision-support pipeline, starting from operational data sources and culminating in a structured data warehouse ready for analytical exploitation.

3.2.2 Synthetic data generation

As no real operational database was directly available for this project, synthetic data were generated to simulate the internal business activity of Bangladesh. The main objective of this step was to create datasets that are both realistic and sufficiently rich to support meaningful analysis within a data warehouse.

The generated data were designed based on publicly available statistical references and structures inspired by data collection forms used by national institutions. Care was taken to ensure logical consistency between entities, temporal coherence, and realistic value ranges.

Enterprises were assigned realistic names inspired by existing companies operating in Bangladesh, which contributes to the credibility of the dataset and facilitates interpretation during analysis. The volume of generated data was deliberately chosen to reflect non-trivial workloads, similar to those encountered in real decision-support systems.

The generated CSV files cover several aspects of internal business activity :

— enterprise-related data (enterprise code, enterprise name, activity sector, enterprise size, market type, region, and operational status),
— internal economic activity data (domestic turnover, internal trade volume, activity dates),
— reference data required to populate the future dimension tables.

These datasets are intentionally **OLTP-oriented**, meaning that they describe elementary business events and transactions. This design choice allows the simulation of realistic source systems from which analytical data are later derived.

3.2.3 Adopted ETL architecture

The ETL architecture implemented in this project is based on a clear separation of data responsibilities and processing stages. Two main databases were created to support this architecture :

— a **staging database**, used as an intermediate storage area for raw data extracted from CSV files,
— a **Data Warehouse database**, structured according to a multidimensional modeling approach.

This separation ensures data traceability and allows transformations to be performed in a controlled environment without altering the original source data. It also simplifies debugging, validation, and future enhancements.

Each ETL process is implemented within a dedicated SSIS package. This modular approach improves readability, facilitates maintenance, and allows individual components of the pipeline to be tested and executed independently.

3.2.4 Loading the staging area

The first step of the ETL process consists of loading the CSV source files into the staging database. For each category of data, a dedicated SSIS package was developed to ensure a clear separation between different data domains.

Each staging load package follows a simple and controlled data flow :

— a *Flat File Source* component to read the CSV file,
— an *OLE DB Destination* component to insert data into the corresponding staging table.

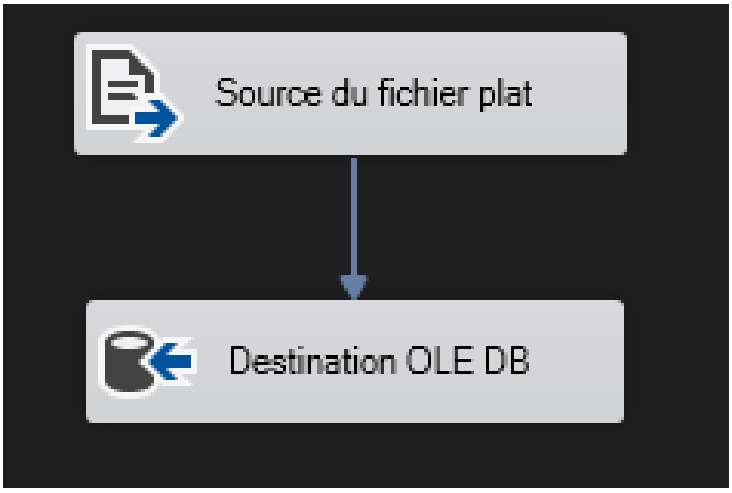

FIGURE 3.9 – Example of an SSIS data flow for loading data into the staging area

Particular attention was paid to data type consistency during this phase, especially for string attributes. A uniform use of Unicode data types was adopted in order to prevent metadata conflicts during joins and transformations in later stages of the ETL pipeline.

3.2.5 Design and implementation of the data warehouse

The data warehouse was designed using a **star schema**, which is well suited for analytical workloads and simplifies query formulation. The central component of this schema is the fact table `FACT_InternalBusiness`, which stores aggregated measures related to internal business activity.

Several dimension tables were created to provide multiple analytical perspectives :
— Time dimension,
— Region dimension,
— Sector dimension,
— Enterprise Size dimension,
— Market Type dimension,
— Enterprise Status dimension.

Each dimension captures descriptive attributes that allow the analysis of facts along different axes. Dimension tables are populated independently from the fact table, using dedicated SSIS packages that extract data from staging tables and apply the necessary transformations.

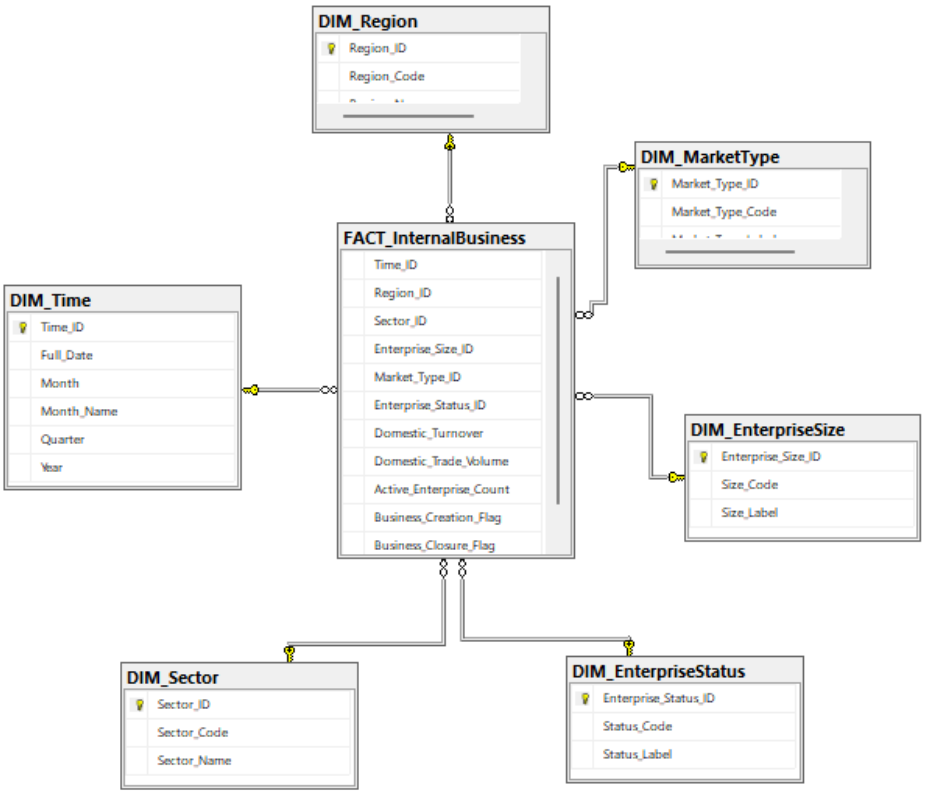

FIGURE 3.10 – Star schema of the Internal Business data warehouse

3.2.6 Fact table loading

The loading of the fact table represents the core of the implementation phase. A dedicated SSIS package was developed to transform operational-level data into aggregated indicators suitable for decision-making.

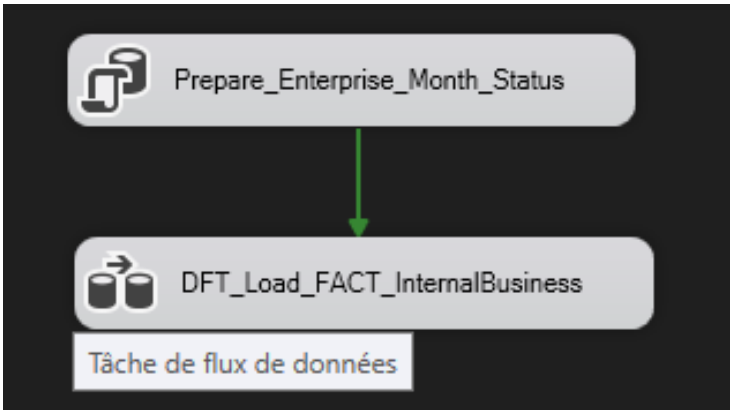

FIGURE 3.11 – SSIS data flow for loading the Internal Business fact table

This package performs a sequence of transformations :

— joining internal activity data with the different dimensions using *Lookup* transformations,

— replacing business identifiers with surrogate keys from dimension tables,

— handling non-matching records through reject flows to ensure data quality,
— aggregating data according to the predefined analytical grain.

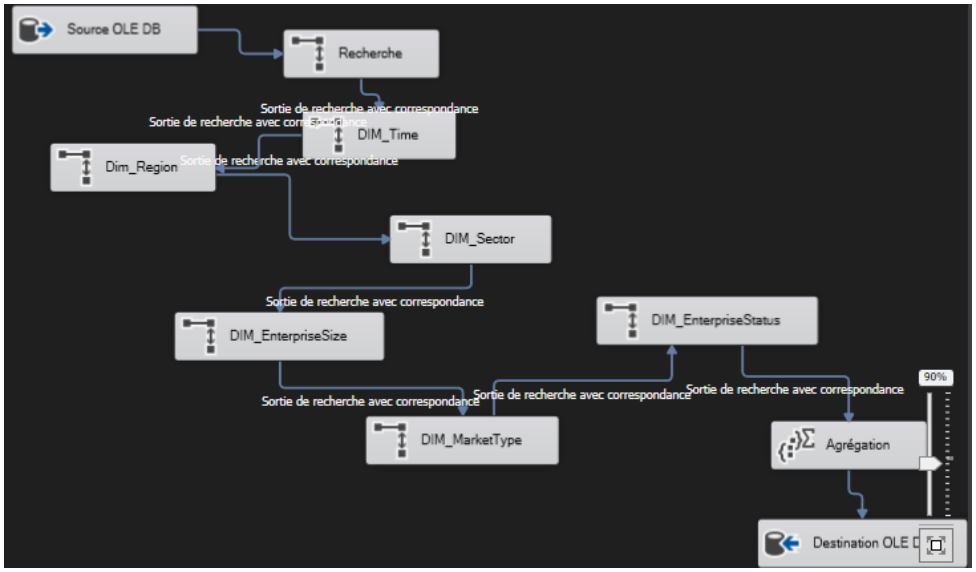

FIGURE 3.12 – Detailed SSIS data flow for fact table processing

The selected grain corresponds to a monthly analysis of internal business activity by :
— time,
— region,
— sector,
— enterprise size,
— market type,
— enterprise status.

Based on this grain, several key indicators are computed :
— domestic turnover,
— internal trade volume,
— number of active enterprises,
— number of business creations,
— number of business closures.

These indicators are calculated using the *Aggregate* transformation before being inserted into the fact table.

3.2.7 Validation and execution of processes

Each SSIS package was tested independently to ensure the correctness of data loading, the validity of joins with dimension tables, and the absence of metadata inconsistencies.

The complete execution of the ETL pipeline results in a fully populated data warehouse that is consistent, reliable, and reproducible. The structured data obtained through this process provide a solid foundation for subsequent analytical queries, reporting, and decision-support visualizations.

```
SELECT TOP 10 * FROM FACT_InternalBusiness;
```

114 %

Résultats | Messages

	Time_ID	Region_ID	Sector_ID	Enterprise_Size_ID	Market_Type_ID	Enterprise_Status_ID	Domestic_Turnover	Domestic_Trade_Volume	Active_Enterprise_Count	Business_Creation_Flag	Business_Closure_Flag
1	1	1	5	1	1	2	17121970	8797	1	0	0
2	2	1	5	1	1	2	15642813	10024	1	0	0
3	3	1	5	1	1	2	20567025	17020	1	0	0
4	4	1	5	1	1	2	18156669	52511	1	0	0
5	5	1	5	1	1	2	19084439	10258	1	0	0
6	6	1	5	1	1	2	16890121	11646	1	0	0
7	7	1	5	1	1	2	18301630	16469	1	0	0
8	8	1	5	1	1	2	18754192	9656	1	0	0
9	9	1	5	1	1	2	14061769	10446	1	0	0
10	10	1	5	1	1	2	21015772	14555	1	0	0

FIGURE 3.13 – Successful execution of the ETL process and fact table loading

3.2.8 Power BI Dashboards

The dashboards are designed to support monitoring and decision-making related to domestic economic activity across regions, sectors, and enterprise characteristics.

Key Performance Indicators

The application presents a set of key performance indicators that provide an overview of internal business activity in Bangladesh :

— Domestic Trade Volume
— Domestic Turnover
— Number of Active Enterprises
— Business Creations
— Business Closures

These indicators allow decision-makers to quickly assess the level of domestic economic activity and enterprise dynamics.

Analytical Visualizations

The Power BI dashboards include several interactive visualizations enabling in-depth analysis :

— Evolution of domestic turnover over time.
— Geographical distribution of domestic trade volume by region.
— Comparison between business creations and closures by sector.
— Top active economic sectors.
— Enterprise status analysis (active vs. closed enterprises).
— Contribution to domestic turnover by enterprise size.

Interactive filters such as *Region, Sector, Enterprise Size, Market Type, and Date* allow users to explore the data from multiple analytical perspectives and customize the analysis according to their needs.

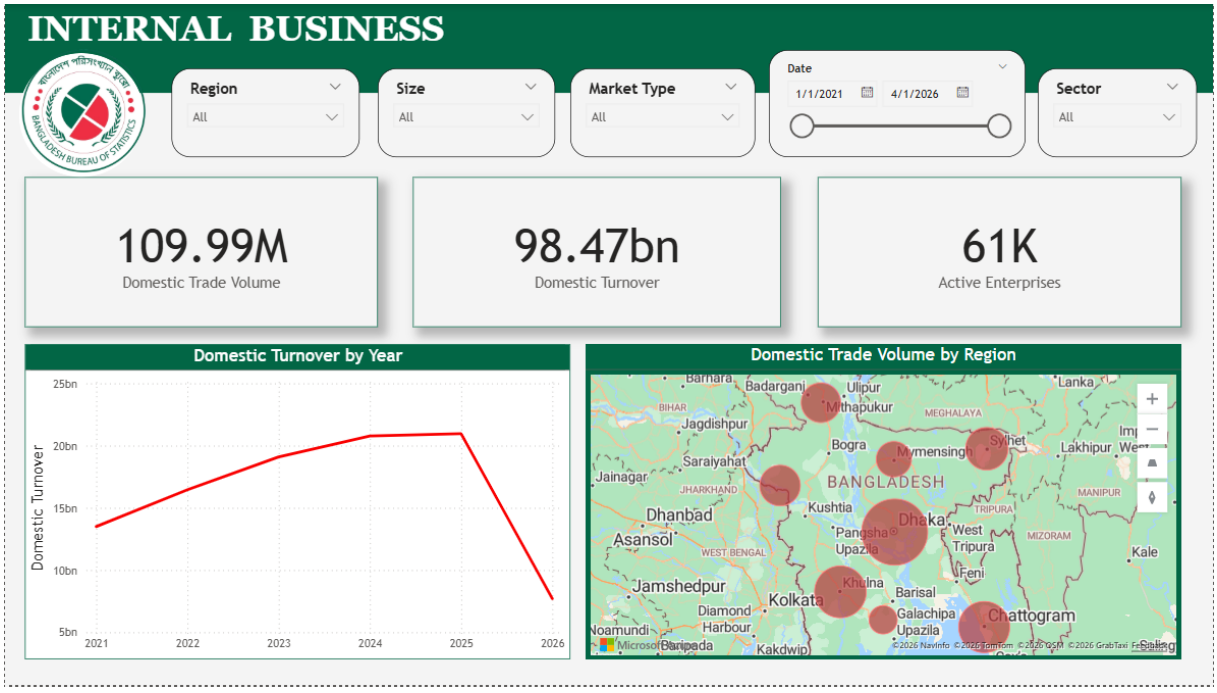

FIGURE 3.14 – Internal Business and Industry Dashboard 1

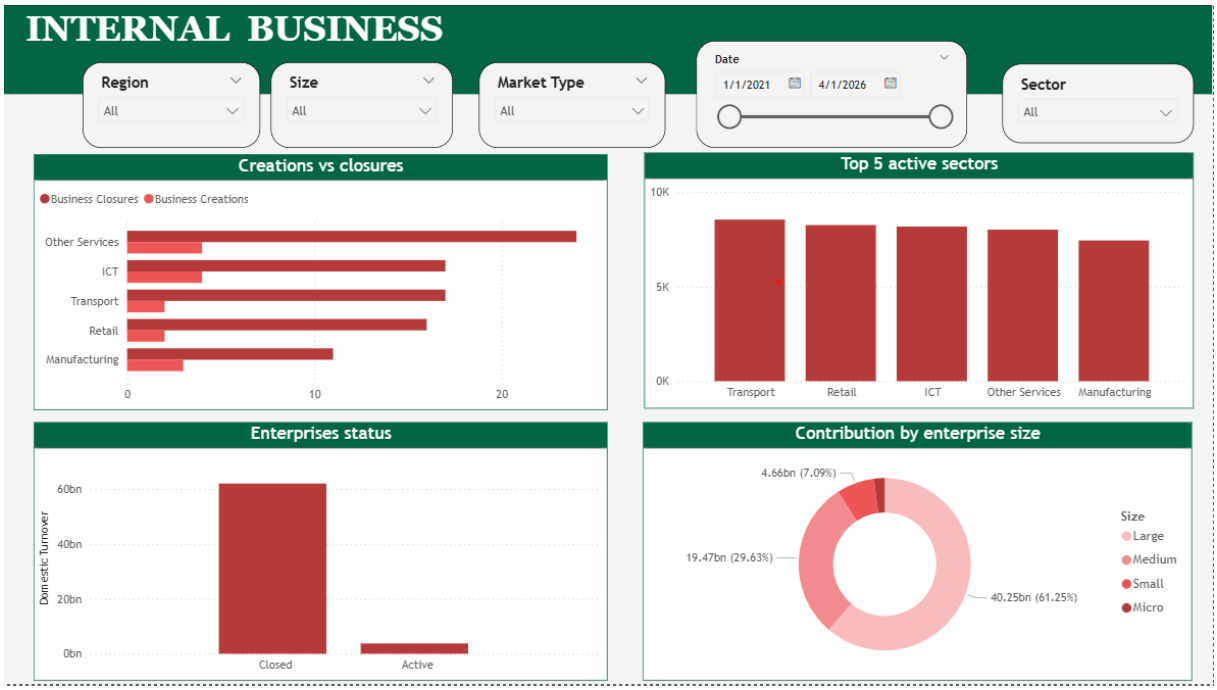

FIGURE 3.15 – Internal Business and Industry Dashboard 2

3.3 External Business

3.3.1 DataMart Physical Design

The External Trade DataMart was designed following a **star schema architecture**, which is well suited for analytical and reporting purposes. This architecture ensures simplicity, high query performance, and ease of use for analytical users.

Dimension Tables

The following dimension tables were implemented in Microsoft SQL Server :
— **DimDate** : provides the temporal analysis axis, including year information.
— **DimCountry** : represents trading partner countries.
— **DimAccount** : defines the type of trade account (Imports or Exports).
— **DimRoute** : indicates the transportation mode (Sea, Land, Air).
— **DimComposition** : describes the type of traded goods (Consumer goods, Capital goods, Materials, etc.).

Each dimension table uses a surrogate key as its primary key, referenced by the fact table to ensure referential integrity.

Fact Table

The central fact table, `FactTrade`, stores quantitative measures related to external trade activities. The main measures include :
— Export value
— Import value
— Trade balance

The fact table references all dimension tables through foreign keys, allowing multidimensional analysis of trade data.

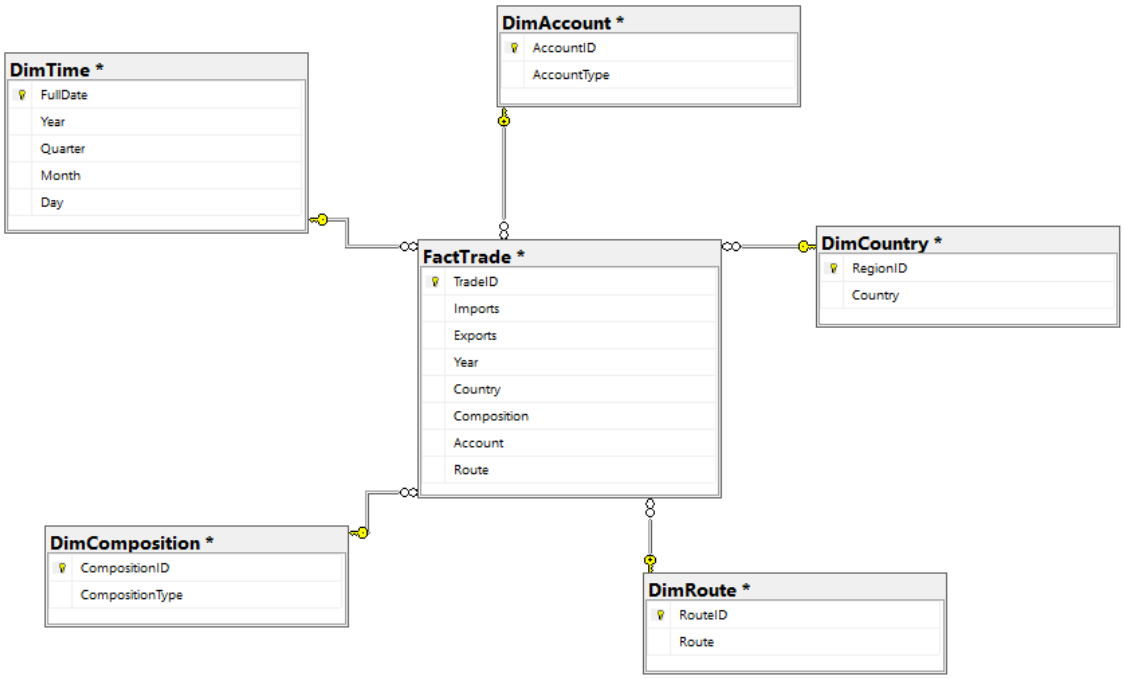

FIGURE 3.16 – Physical DataMart in SQL Server

3.3.2 ETL Conception

The ETL processes were designed to integrate data from heterogeneous sources into the Data Warehouse in a reliable and structured manner. The source data consists of **Excel files containing yearly survey data** on Bangladesh's external trade sector.

ETL Architecture

The ETL architecture is based on a clear separation between dimension loading processes and fact loading processes. Dimension tables are populated first in order to provide reference data required by the fact table.

ETL Design for Dimension Tables

For each dimension, a dedicated ETL data flow was designed. The design principles applied to dimension ETL processes are as follows :
— Extraction of data from Excel survey files.
— Selection of descriptive attributes relevant to each dimension.
— Identification and elimination of duplicate records.
— Preparation of clean and consistent reference data for loading.
This design ensures the uniqueness and consistency of dimension data.

ETL Design for the Fact Table

The ETL design for the fact table is based on the use of descriptive attributes present in the source data and their transformation into foreign keys.

The design includes :

— Mapping of source attributes to dimension tables.
— Definition of lookup relationships between fact data and dimensions.
— Preparation of quantitative measures for analytical use.

3.3.3 ETL Development Using SSIS

The ETL processes were implemented using **SQL Server Integration Services (SSIS)**, in accordance with the previously defined ETL design.

Development of Dimension ETL Data Flows

Each dimension ETL data flow was developed following the same implementation logic :

1. Connection to Excel files containing yearly survey data.
2. Selection of the columns required for the dimension.
3. Use of a *Sort* transformation with the *Remove Duplicates* option to ensure uniqueness.
4. Loading of distinct records into the corresponding dimension table.

This implementation guarantees clean and non-redundant dimension data.

Development of the Fact Table ETL Data Flow

The ETL data flow for the `FactTrade` table was developed to integrate transactional survey data into the Data Warehouse.

The implementation steps are as follows :

1. Extraction of trade data from Excel survey files.
2. Application of multiple *Lookup* transformations to retrieve surrogate keys from each dimension table.
3. Replacement of descriptive attributes with foreign keys.
4. Loading of the transformed records into the `FactTrade` table.

This approach ensures data consistency and referential integrity.

ETL

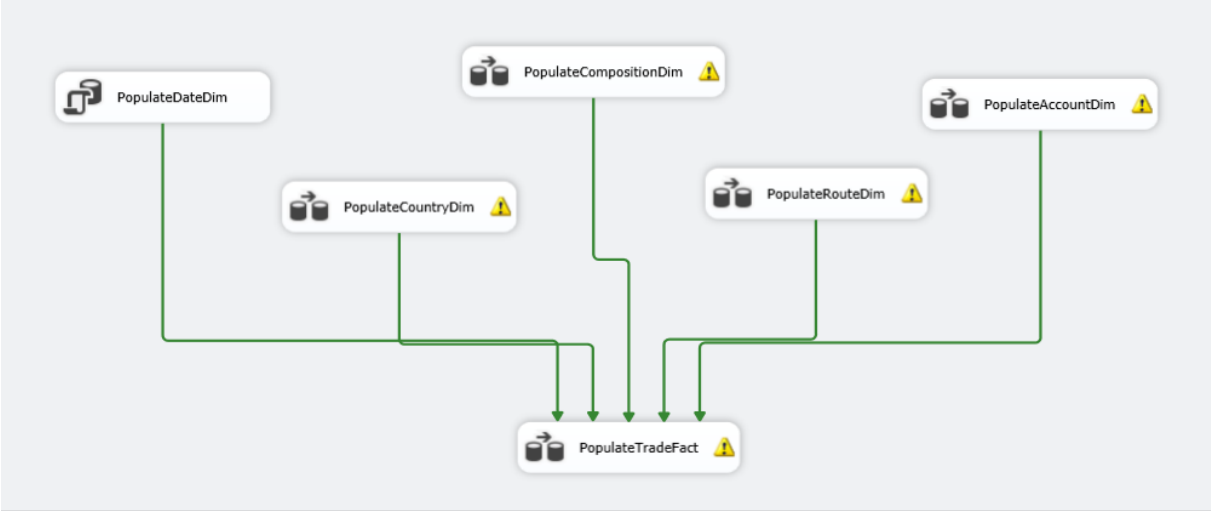

FIGURE 3.17 – External Business ETL in SSIS

3.3.4 User Application : Power BI Dashboards

The user application was developed using **Microsoft Power BI** and provides an interactive interface for analyzing external trade data stored in the Data Warehouse.

Key Performance Indicators

The application displays several key performance indicators that summarize the state of Bangladesh's external trade :
— Total Imports
— Total Exports
— Imports Growth
— Exports Growth
— Trade Balance
— Comparison with the previous year
These indicators provide a quick overview for decision-makers and analysts.

Analytical Visualizations

The Power BI application includes interactive visualizations such as :
— Imports and Exports by country.
— Trade values by goods composition type.
— Trade analysis by transportation route.
— Yearly evolution of imports, exports, and trade balance.
Dynamic filters (Year, Account Type, Country, Route, Composition) allow users to explore the data from multiple analytical perspectives.

Dashboard

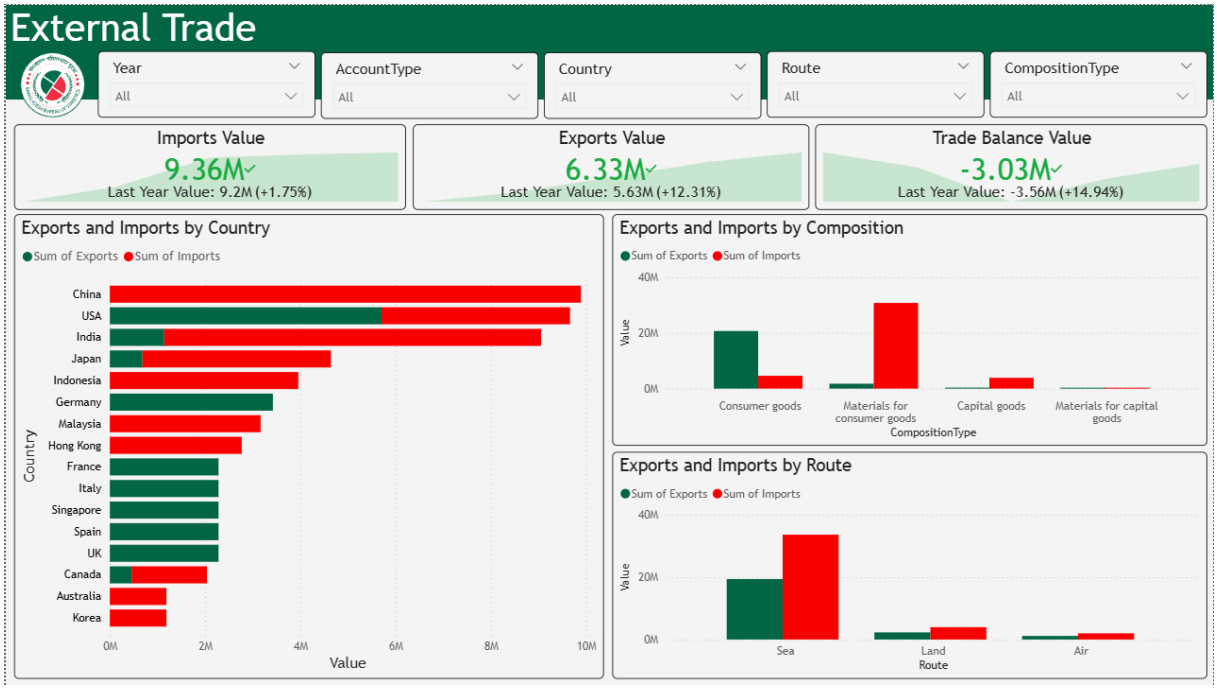

FIGURE 3.18 – External Business Dashboard in Power BI

Conclusion

This project successfully delivered a comprehensive Business Intelligence solution for the Industry & Business domain of the Bangladesh Bureau of Statistics. Through the implementation of three integrated datamarts, Industry, Internal Business, and External Business, we established a robust analytical platform supporting evidence-based decision-making for economic policy and national development planning.

The technical implementation demonstrates effective application of dimensional modeling, ETL best practices, and modern BI technologies. Star schema architectures ensure optimal query performance, while SSIS-based ETL processes provide reliable data integration from heterogeneous sources. Interactive Power BI dashboards translate complex statistical data into actionable insights accessible to policymakers, analysts, and the public.

Key achievements include the successful integration of industrial performance indicators, domestic market metrics, and international trade statistics into a unified analytical framework. The system enables multidimensional analysis across temporal, geographic, sectoral, and categorical dimensions, supporting comprehensive monitoring of Bangladesh's economic activity.

The modular architecture and documented methodologies provide a replicable model for extending the statistical data warehouse to additional domains. Future enhancements may include real-time data integration, predictive analytics capabilities, and expanded geographic analysis through GIS integration.

This work contributes to strengthening Bangladesh's statistical infrastructure, demonstrating how Business Intelligence technologies can transform fragmented data into strategic intelligence that supports national development objectives and international reporting commitments.