Politechnika Wrocławska

Centrum Wiedzy i Informacji Naukowo-Technicznej

Wroclaw University of Science and Technology

Faculty of Information and Communication Technology

Application programming - Data mining and data warehousing

Lab 1 - Report

Report by:

Hedi Hassan (283496)

Data Warehousing Lab Report

1. Introduction

This project focuses on building a data warehouse and analytical reporting solution for academic data using Microsoft SQL Server Integration Services (SSIS) and SQL Server Analysis Services (SSAS). The objective was to import, clean, and transform raw data related to students, teachers, grades, and courses; create a multidimensional cube; and provide insights through KPIs and hierarchical dimensions.

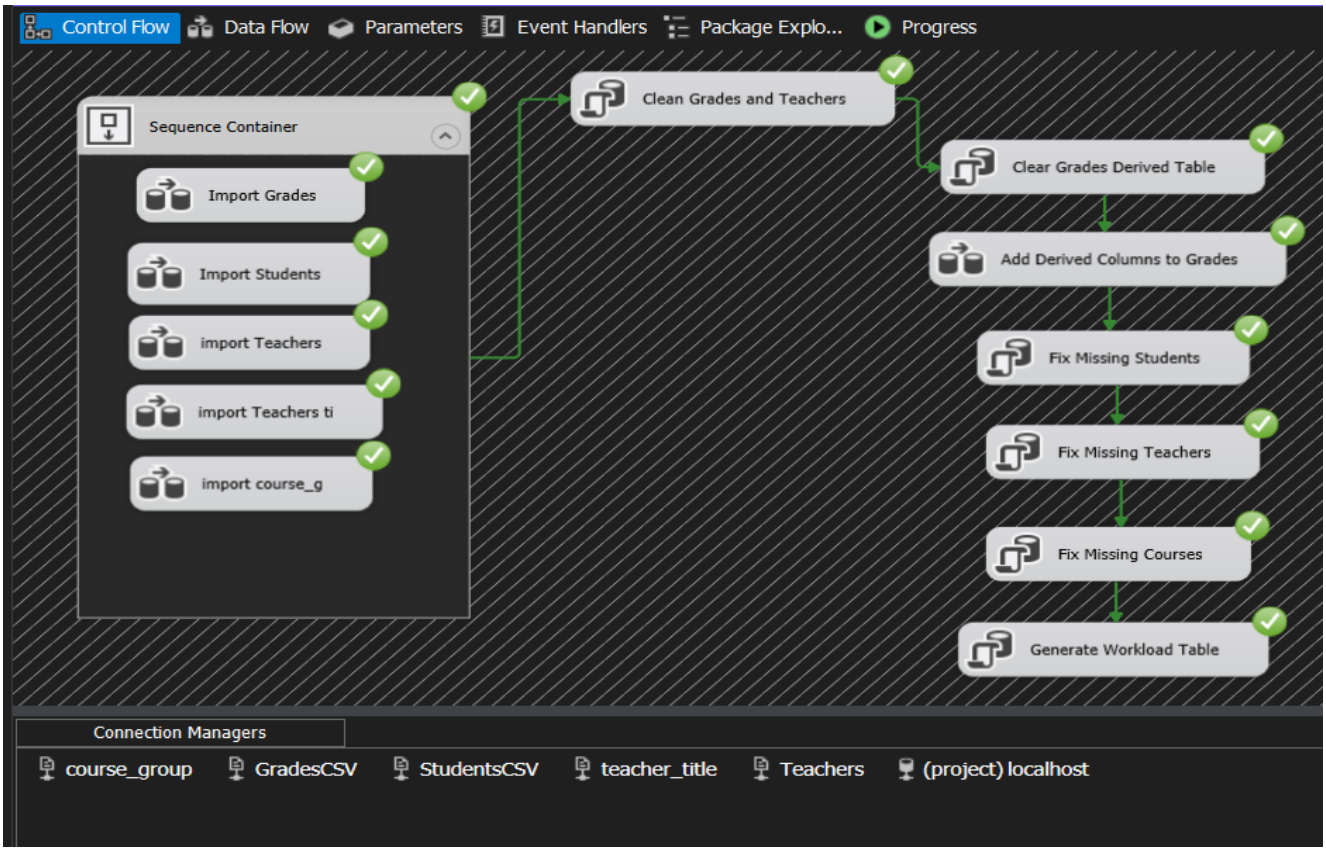

Figure 1: ETL Workflow (SSIS)

2. ETL Workflow (SSIS) The ETL workflow is designed to import, clean, enrich, and transform raw data from multiple CSV files into a structured warehouse format. The control flow consists of the following key steps:

2.1 Data Import

- **Import Grades:** Loads raw grade records from CSV into the grades table.
- **Import Students:** Loads student demographic and academic data into the students table.
- **Import Teachers:** Loads teacher data including faculty, title ID, and institute.
- **Import Teacher Titles:** Imports title information (e.g., Assistant, Professor) into teacher_title.
- **Import Course Groups:** Loads mapping of course codes to course group categories.

2.2 Data Cleaning and Preparation

- **Clean Grades and Teachers:** Removes illegal grades, fills missing fields with default values, and standardizes teacher gender and titles.
- **Clear Grades Derived Table:** Ensures grades_derived is emptied before populating fresh records.

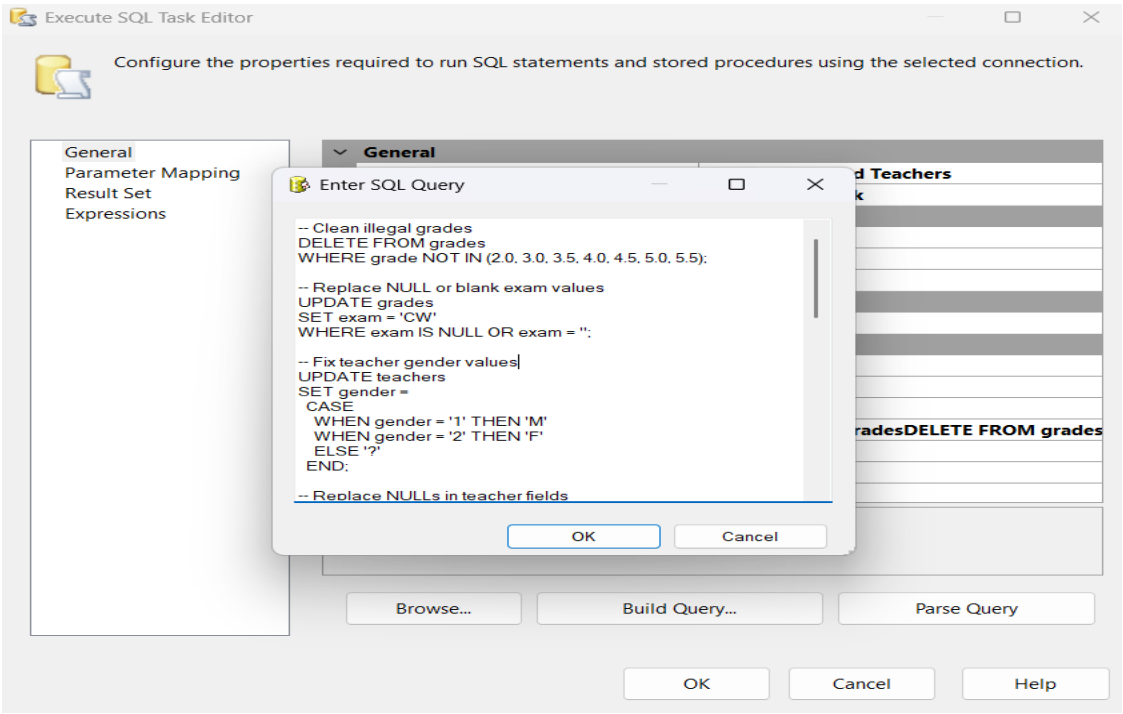

Figure 2: Clean Grades and Teachers

- **Add Derived Columns to Grades:** Enriches grade data with derived attributes such as semester type, academic year, and normalized grade calculations.

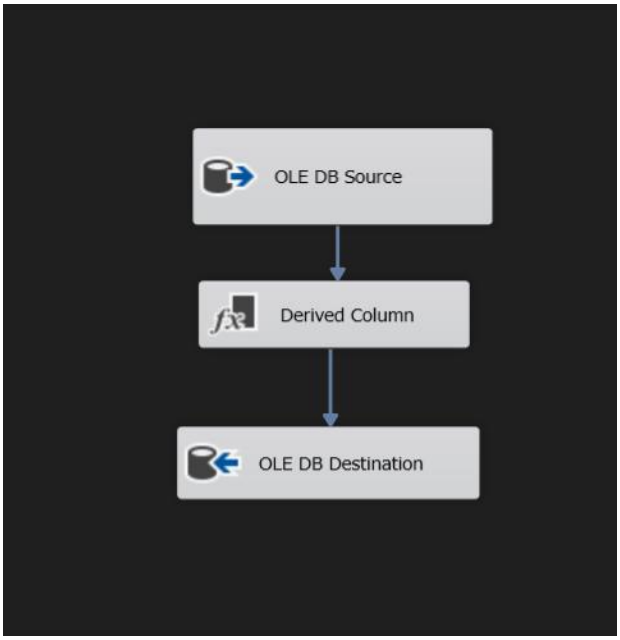

Figure 3: Add Derived Columns to Grades

2.3 Fix Missing Dimension Entries

- **Fix Missing Students:** Adds any students referenced in grades but missing in the students table.

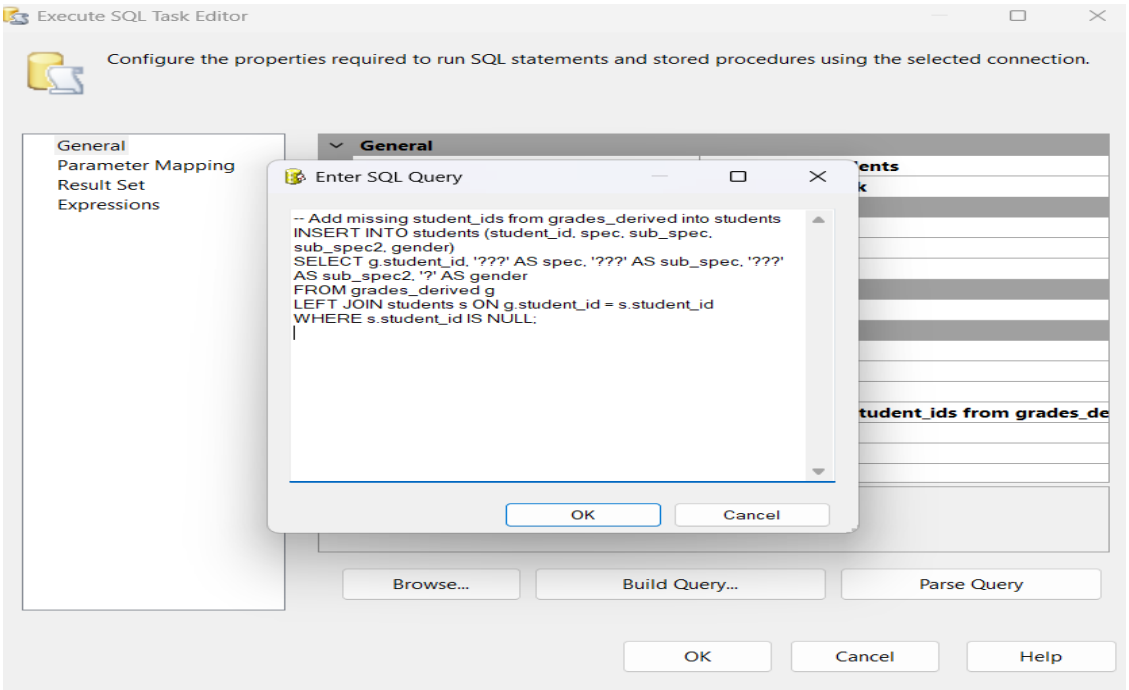

Figure 4: Fix Missing Students

- **Fix Missing Teachers:** Adds placeholder entries for teachers found in grades_derived but absent in teachers.

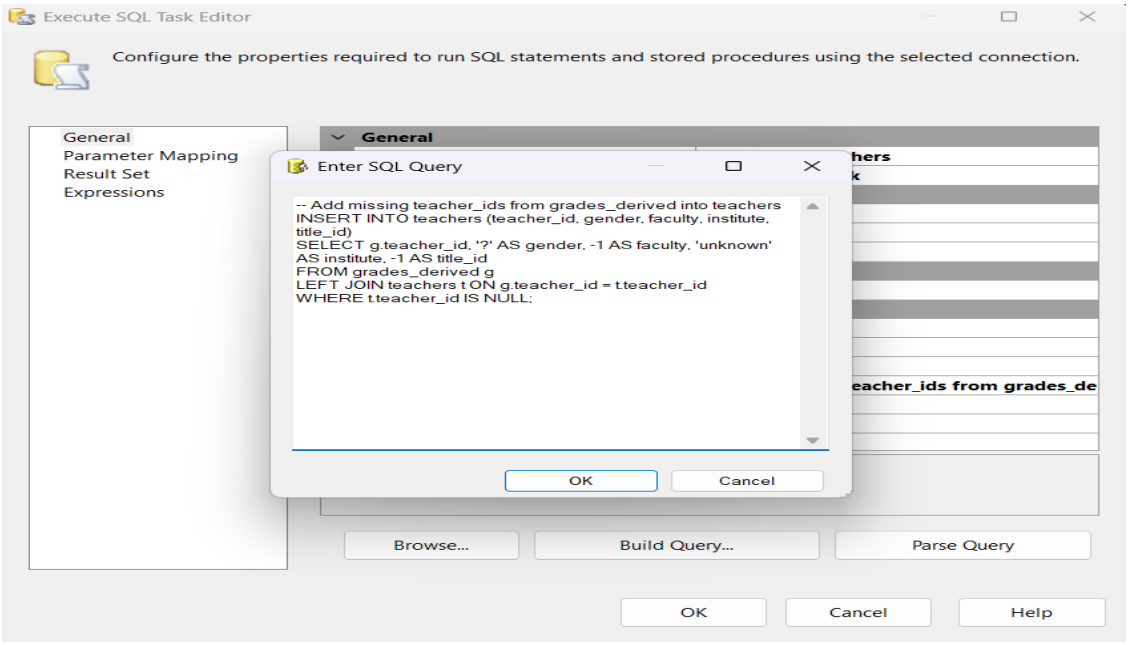

Figure 5: Fix Missing Teachers

- **Fix Missing Courses:** Inserts unknown course entries referenced in grades but missing from course_group.

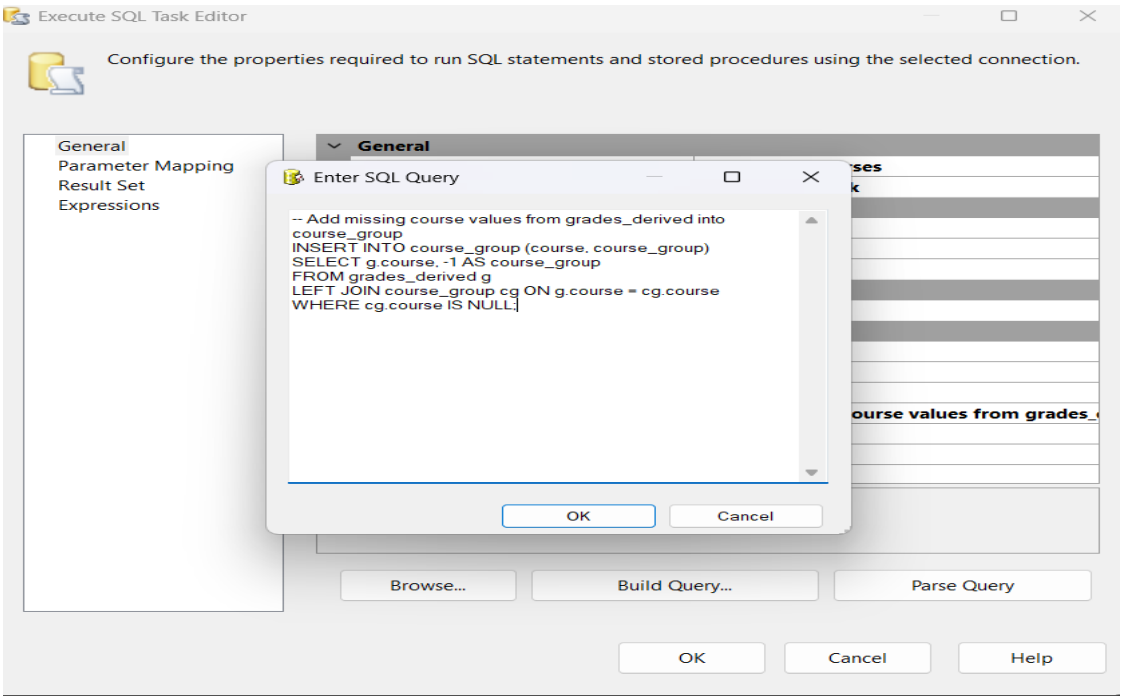

Figure 6: Fix Missing Courses

2.4 Generate Analytical Tables

- **Generate Workload Table:** Computes the number of students per teacher and stores it in a separate table for workload analysis.

3. Data Warehouse Design (SSAS)

An OLAP cube (DWLab_Cube2) was developed using the transformed data. The cube includes the following dimensions:

- **Students:** Gender, Spec, Sub Spec, Sub Spec2
- **Teachers:** Faculty, Institute, Title, Title ID
- **Course Group:** Course, Course Group
- **Academic Year** and **Semester Type:** Derived from the grade date for time-based analysis
- **Teacher Workload:** Includes composite keys and workload facts

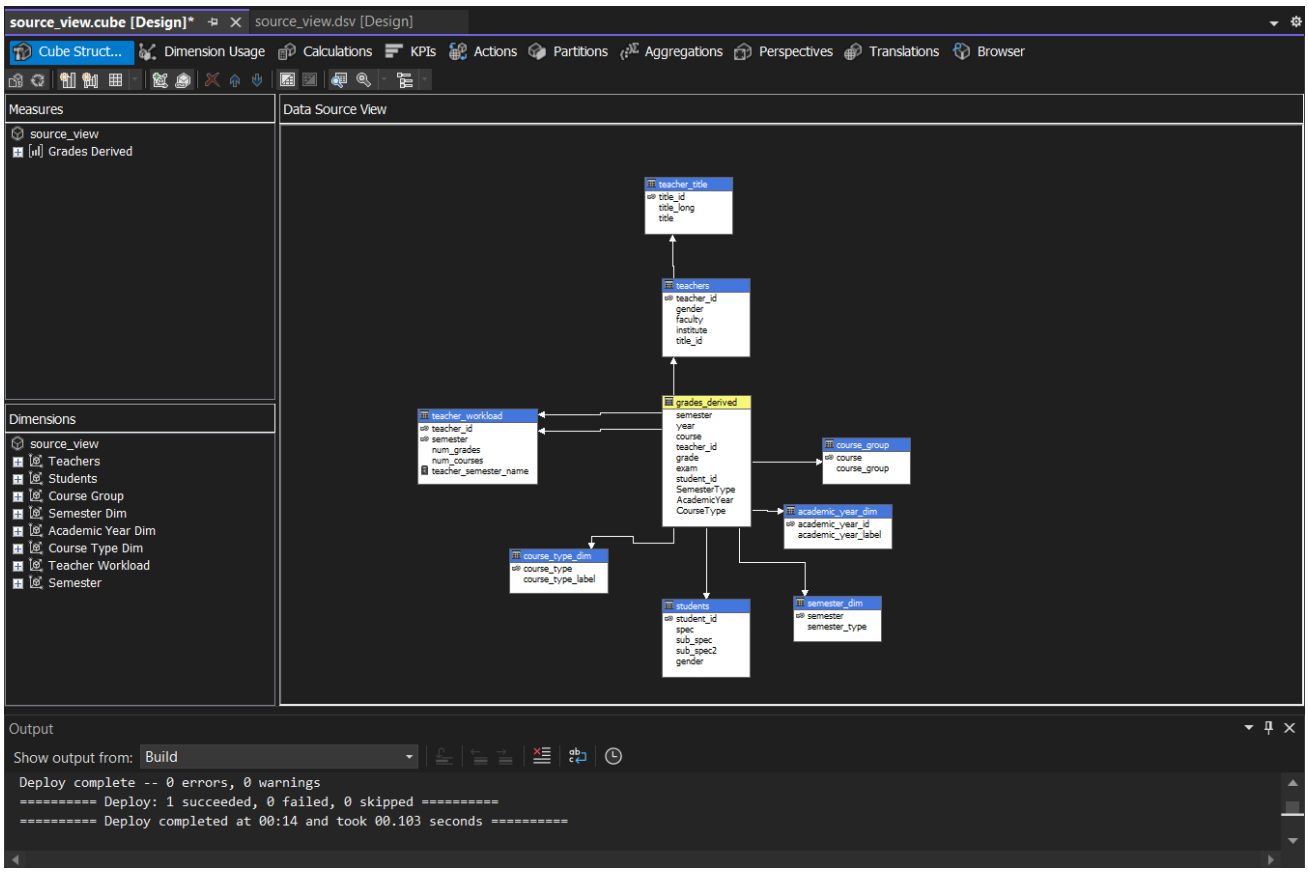

Figure 7: Cube Structure

4. KPIs and Hierarchies

- A KPI named **"Student Grade KPI"** was created to compare average grades by gender against a fixed goal (e.g., 4.0).

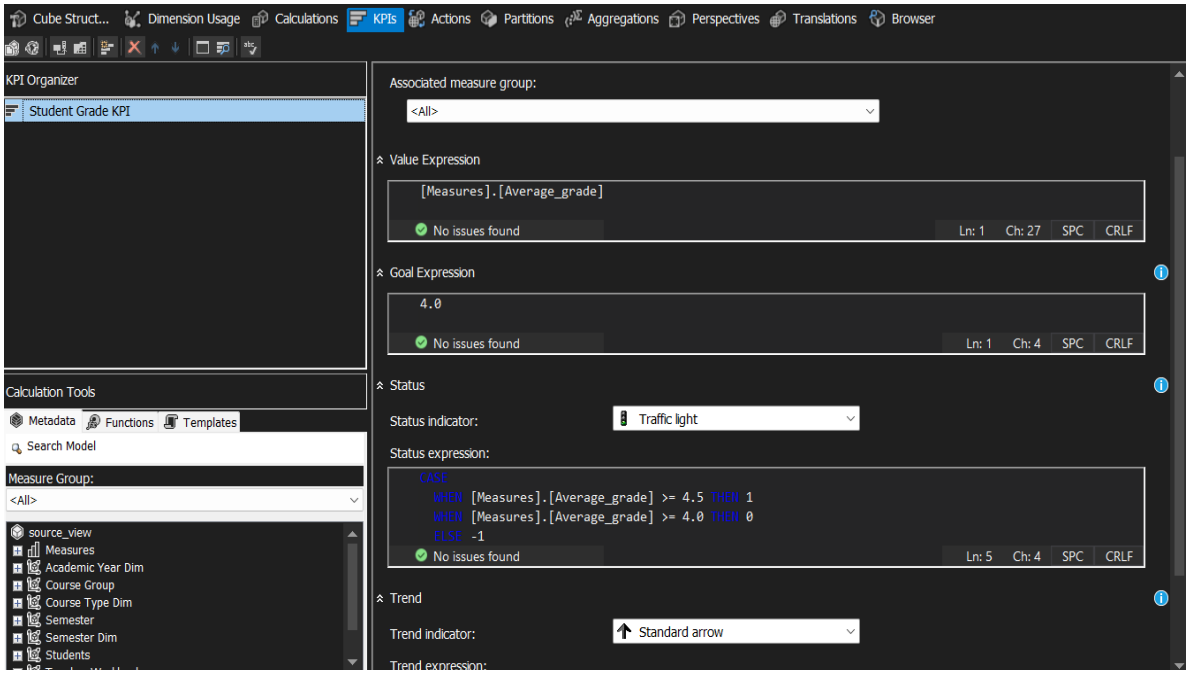

Figure 8: Student Grade KPI Creation

- Hierarchies were established within the student dimension to enable drill-down from Spec → Sub Spec → Sub Spec2.

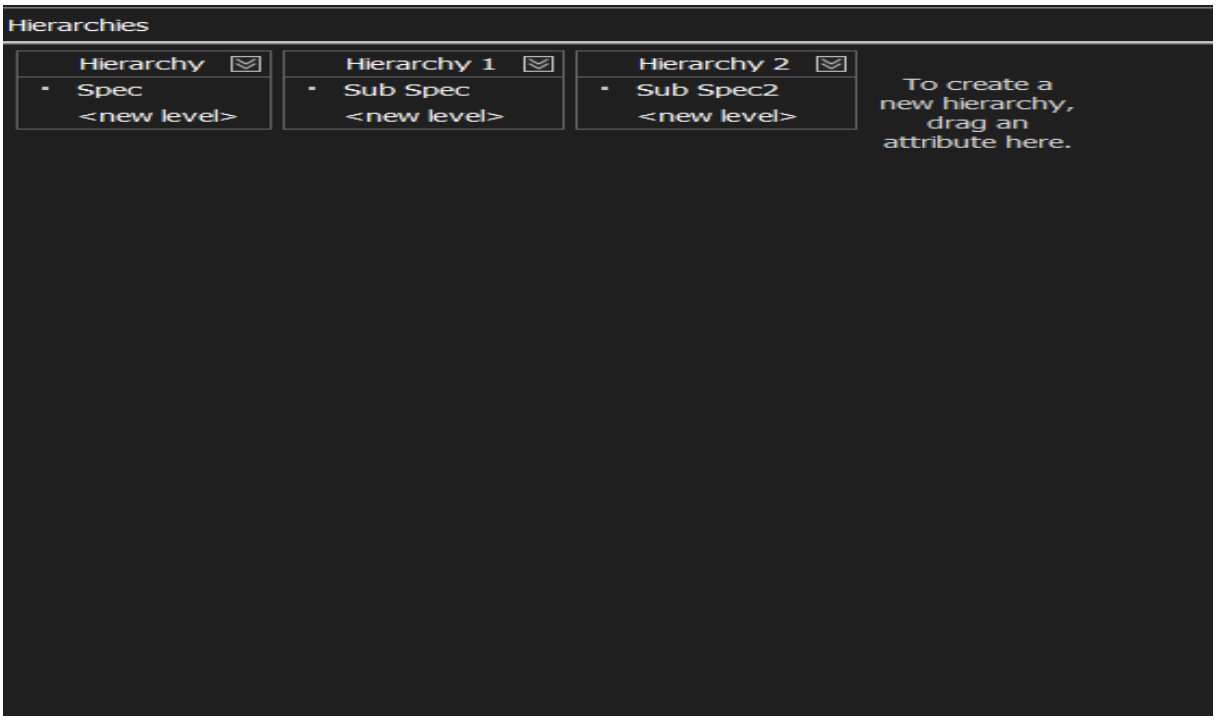

Figure 9: Hierarchies

5. Results and Insights

Sample pivot tables and KPIs in the browser confirmed correct aggregations:

- A multidimensional analysis was performed to display the average student grade grouped by academic year and course type. The results showed that average grades vary by course type. The data includes Exercise, Lab, Lecture, Project, and Seminar types.

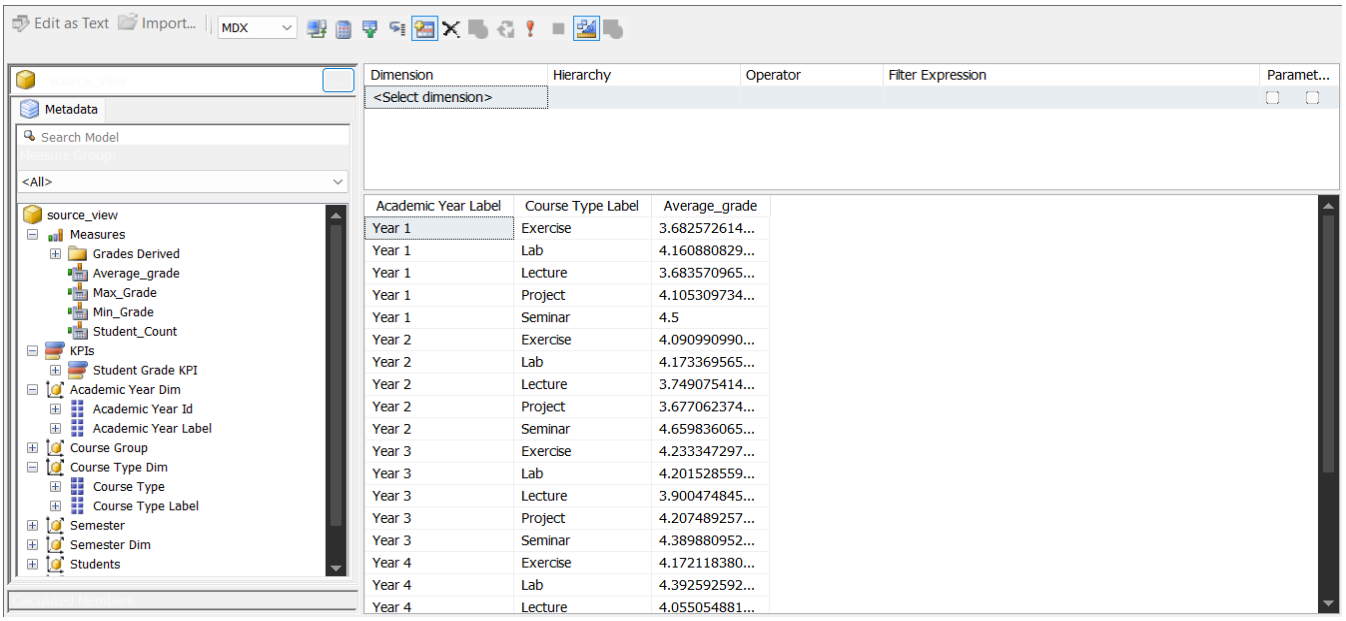

Figure 10: Average Grade by Academic Year and Course Type derived from multidimensional cube analysis.

- A multidimensional analysis was executed using the hierarchy of Gender and Spec to evaluate the Average Grade measure. The pivot table displays average grade values segmented by gender and academic specialization.

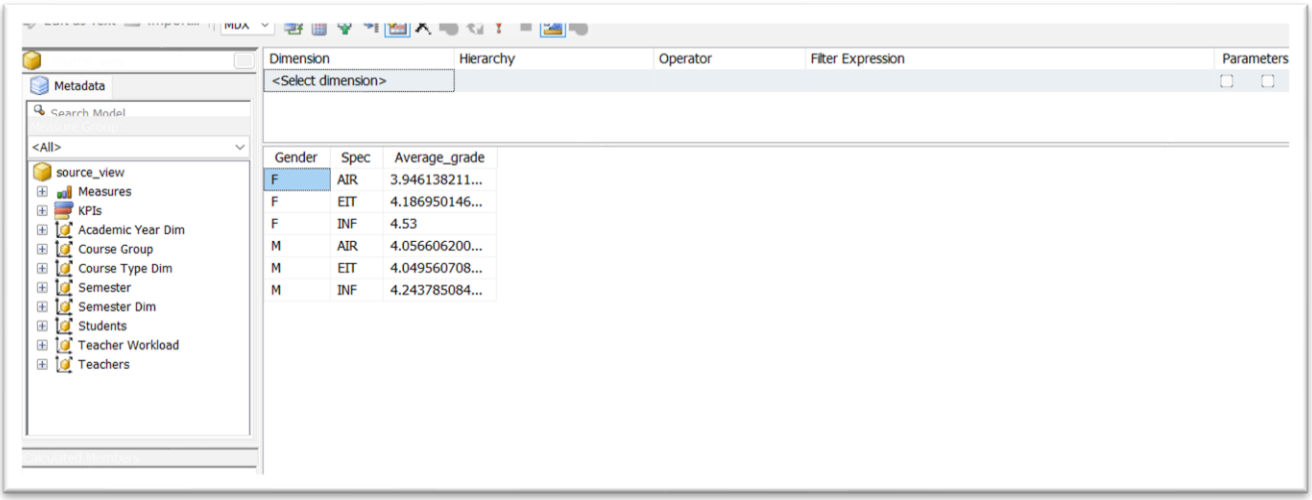

Figure 11: Average Grade segmented by Gender and Specialization (Spec) using a hierarchical dimension drill-down.

6. Data Source View and Cube Structure

The final Data Source View (DSV) defines all the relational links among key tables: grades_derived, students, teachers, teacher_title, course_group, teacher_workload, and additional dimensions like academic_year_dim and semester_dim. These relationships enable the SSAS cube to perform accurate aggregations and queries.

The Cube Structure binds the Grades Derived measure group to these dimensions. Deployment was successful, indicating a validated design. Inside the cube, calculated measures such as Average_grade, Min_Grade, Max_Grade, and Student_Count were created to enhance analytical capabilities.

7. Cube Browser Configuration

In the cube browser, users can perform ad-hoc exploration by dragging dimensions and measures into the query pane. The cube exposes KPIs and hierarchies that enable intuitive filtering and slicing, supporting insights like grade trends per academic year, student performance by spec, and semester comparisons.

8. Conclusion

The ETL and cube development process successfully enabled advanced analytics on student performance and teacher workload. All required dimensions were built, data was validated, and MDX queries provided meaningful insights. The project now supports slicing and dicing by academic year, semester, course group, and more.