

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

## Term Project – Fall 25

### OBJECTIVE

This project aims to equip students with complete hands-on experience across the **entire Business Intelligence lifecycle**. Students will extract data from **multiple OLTP sources**, perform **ETL transformations** in SSIS, design and populate a **Dimensional Data Warehouse**, build an **OLAP Cube in SSAS**, and finally develop a **Business Intelligence Dashboard in Power BI** for analytical insights.

The project mirrors real-world BI development where organizations consolidate data from different transactional systems and external datasets, clean & harmonize inconsistencies, and enable powerful decision-making through OLAP and dashboarding.

### LEARNING OUTCOMES

After completing this project, students will be able to:

- Design a real-world multi-source ETL pipeline
- Resolve data inconsistencies across heterogeneous systems
- Implement staging, transformations, and dimension/fact loading
- Generate a continuous Time dimension
- Build an OLAP cube and perform BI operations (roll-up, drill-down, slice, dice)
- Conduct analytical reasoning using cube measures
- Develop corporate dashboards with advanced insights in Power BI
- Understand the full BI lifecycle from source systems to decision-making

### CASE STUDY BACKGROUND

Northwind Global Distributors, a global retail distribution company, uses two independent transactional databases, **Northwind\_Source1** and **Northwind\_Source2**, each containing orders, customers, employees, products, categories, suppliers, shippers, order details, and the regional structures Region, Territories, and the EmployeeTerritories junction. Both systems share similar structures but contain differences in data and overlapping records.

Additionally, the Marketing department provides **two external files** containing Customer Contact Names and Genders. Each file encodes gender differently:

File 1 → "M" and "F"

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

File 2 → "1" (Male) and "0" (Female).

Your job is to integrate these three heterogeneous data sources into a **single, consistent Data Warehouse** whose structure has already been provided to you (DW\_Northwind schema). Before implementing the ETL in SSIS, you must first design a BPMN model for the ETL pipeline. Then, using SSIS, design and implement the full ETL pipeline; using SSAS, create an OLAP cube for analytical exploration; and finally connect the data warehouse to Power BI to produce a corporate analytics dashboard.

## PROJECT TASKS

### TASK 1 — Problem Analysis (Understanding Sources & DW Design)

Analyze the provided data sources:

- Two OLTP databases: **Northwind\_Source1** and **Northwind\_Source2** (creation + insertion scripts)
- Two external files containing **Customer Contact Names and Gender values**

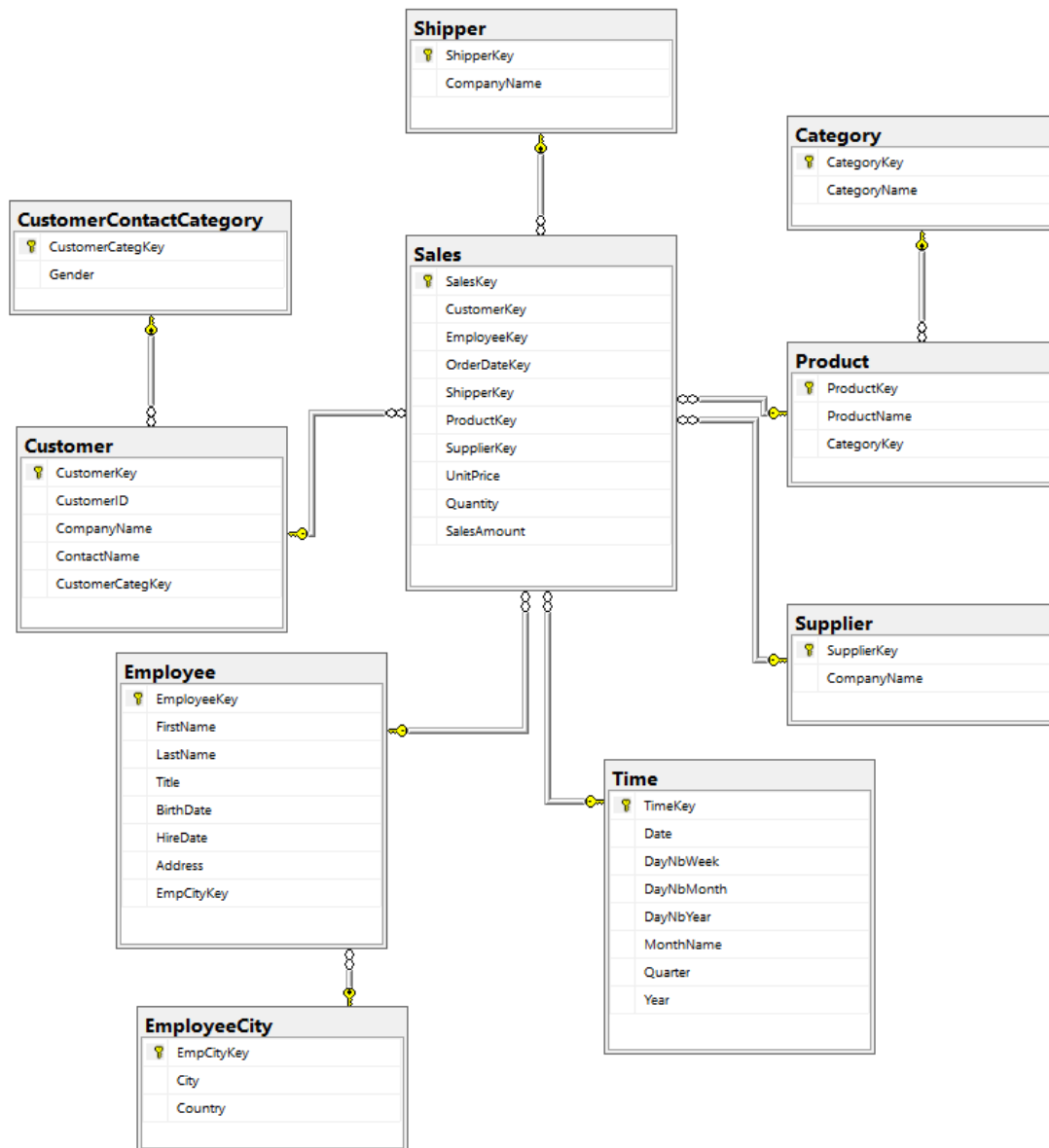
Your analysis should focus on:

- Understanding the overall structure and content of both OLTP databases, including overlaps and differences in customers, employees, products, orders, and related tables
- Identifying general data quality issues such as overlapping records between sources, missing values, and variations in identifiers.
- Understanding the grain of the Sales fact table.
- Understanding the provided **DW\_Northwind** schema, including which tables use **business keys** directly and which tables use **surrogate keys**, as defined in the DW creation script
- Understanding how external gender data relates to customer contact names

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

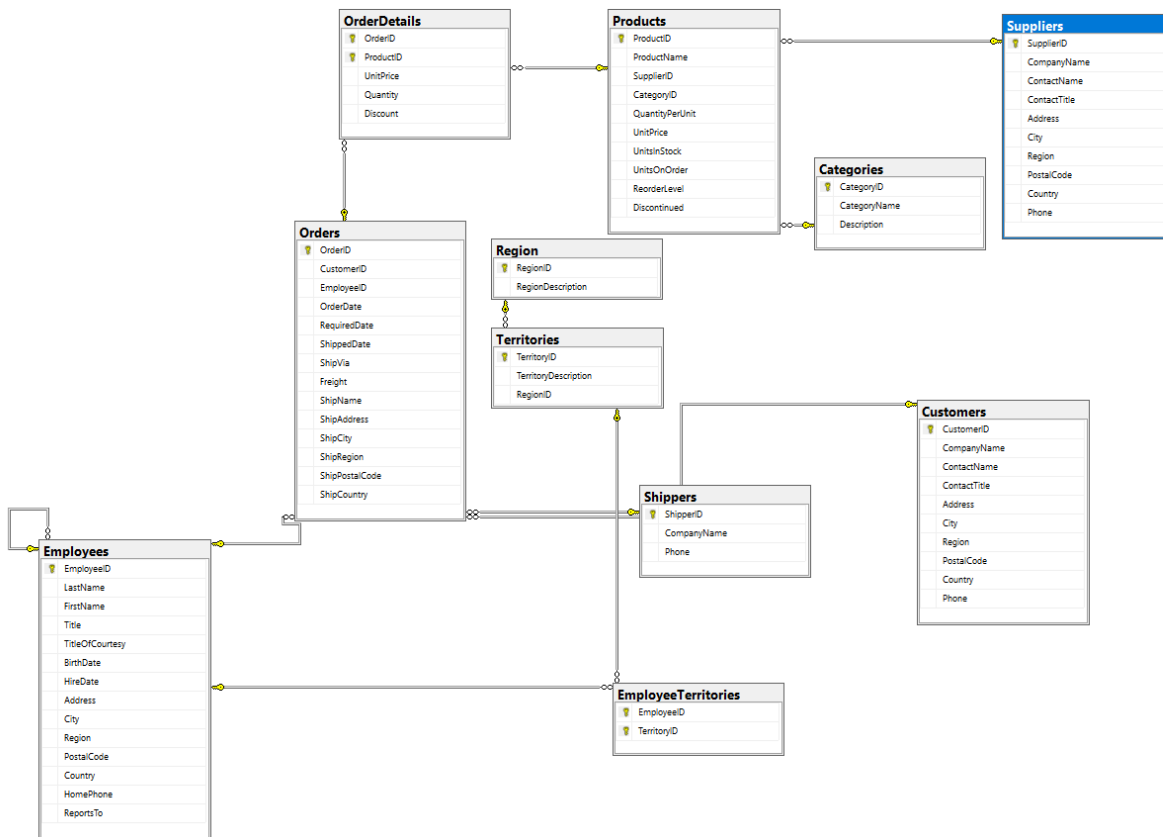


(Fig: Datawarehouse Schema)

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23



(Fig: Databases Schema)

## TASK 2 — Workflow Design (BPMN)

Design BPMN diagrams for the ETL process that you will later implement in SSIS.

### 2.1 Control Flow BPMN

The diagram should reflect the high-level ETL workflow, including:

- Extraction from Northwind\_Source1
- Extraction from Northwind\_Source2
- Extraction from both external gender files

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

- Loading all extracted data into a staging area
- Transformation and integration steps
- Loading dimension tables into the data warehouse
- Loading the Sales fact table

## 2.2 Data Flow BPMN

Design a separate data flow view showing how data moves for major entities:

- Flow from sources → staging → transformations → DW tables
- Customer flow must include gender standardization using external files
- Employee flow must include handling of employee city information
- Product flow must reflect linkage with category data
- Sales flow must combine orders and order details with required dimension keys

The BPMN diagrams should be clear, readable, and aligned with how SSIS packages are typically designed. They should be created using a BPMN tool, such as Camunda.

## TASK 3 — ETL Implementation in SSIS

Implement the BPMN design using SSIS.

### 3.1 Staging Area Load

Create staging tables for all required OLTP tables and external files.

Load data into staging **without applying transformations**.

### 3.2 Transformations

- Standardize gender values from external files into a common representation before loading them into the customer-related dimension level

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

- Resolve overlaps between the two source systems during integration
- Prepare date values required for the Time dimension

## ⇒ Time Dimension Instruction (Required):

Use a stable **TimeKey** format as an integer in **YYYYMMDD** form.

Formula:

$\text{TimeKey} = \text{YEAR}(d) * 10000 + \text{MONTH}(d) * 100 + \text{DAY}(d)$

Populate a **continuous calendar** covering the required date range (for example, from the minimum OrderDate to the maximum OrderDate across both sources).

## ⇒ Key Design Instructions for DW\_Northwind

In the DW\_Northwind schema, both **business keys** and **surrogate keys** are used, as defined below. Students must load and reference tables accordingly during ETL:

### Business keys (taken directly from OLTP sources, NOT identity):

- **Category** → **CategoryKey**
- **Product** → **ProductKey**
- **Supplier** → **SupplierKey**
- **Shipper** → **ShipperKey**
- **Employee** → **EmployeeKey**
- **Time** → **TimeKey** (INT in YYYYMMDD format)

### Surrogate keys (generated in the Data Warehouse using IDENTITY):

- **EmployeeCity** → **EmpCityKey**
- **CustomerContactCategory** → **CustomerCategKey**

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

- **Customer** → **CustomerKey** (while **CustomerID** remains the business key from OLTP and must be unique)
- **Sales (Fact table)** → **SalesKey**

## 3.3 Dimension Loading

Load all dimension tables according to the provided **DW\_Northwind** schema, respecting the defined use of business keys and surrogate keys.

## 3.4 Fact Table Loading (Sales)

Load the Sales fact table at the defined grain.

Calculate SalesAmount during the load and ensure all required dimension references are correctly resolved.

After completing the ETL and loading the data warehouse, create an **SSAS cube** based on the DW schema.

## TASK 4 — SSAS OLAP CUBE DEVELOPMENT

After completing the ETL and loading the data warehouse, create an **SSAS cube** based on the DW schema.

### 4.1 Create Dimensions

Include:

- DimCustomer
- DimEmployee
- DimEmployeeCity
- DimProduct
- DimCategory
- DimSupplier

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

- DimShipper
- DimTime

Define hierarchies:

- **Time Hierarchy:** Year → Quarter → Month → Day
- **Product Hierarchy:** Category → Product
- **Employee Geography:** Country → City → Employee

## 4.2 Create Measures

- Total Sales Amount
- Total Quantity
- Average Unit Price
- Average Discount
- Order Count

## 4.3 Perform OLAP Operations (Required Demonstrations)

### A) Roll-Up

1. Roll up sales from Day → Month → Quarter → Year
2. Roll up Product → Category
3. Roll up EmployeeCity → Country

### B) Drill-Down

1. Year → Quarter → Month → Day
2. Category → Product
3. Country → City → Employee

### C) Slice

1. Sales where Shipper = AsiaCargo
2. Sales where Category = Bakery or Seafood
3. Sales where Gender = Female

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

## D) Dice

1. Sales in 2025 + Category = Seafood + Employee Country = Canada
2. Sales for Products with Quantity > 20 in October
3. Compare two Employees across two Categories

## 4.4 Analytical Questions (Must Answer Using Cube)

1. Which employee generated the highest revenue in 2025?
2. Which category performed best each quarter?
3. Which supplier contributes most revenue within each category?
4. Which shipper handles the most high-value orders?
5. Which product has the highest average order size?
6. Are male or female customers more profitable?
7. Which months show unusual drops in sales?
8. Top 5 customers by revenue.

Include screenshots of each result in your report.

## TASK 5 — POWER BI ANALYTICS DASHBOARD

Connect Power BI to your **SQL Server Data Warehouse (DW)** and import all required DW dimension and fact tables. Model the relationships according to your DW star schema and create the required DAX measures.

Your dashboard **must** include the visuals and analytical components outlined below.

### 5.1 Mandatory Visuals

1. **Time-Series Sales Trend (Year → Quarter → Month)**
  - Use DimTime hierarchy.
  - Must show comparison across 2024 and 2025.
2. **Sales by Category**
  - Drillable to Product level.
3. **Sales by Product**

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

- Ranked bar chart or treemap.
- 4. **Sales by Supplier**
  - Show supplier contribution across both sources.
- 5. **Sales by Shipper**
  - Compare freight cost vs total sales delivered.
- 6. **Employee Sales by Country → City**
  - Geography hierarchy required: Region → Country → City.
  - Use DimEmployee + DimTerritory + FactSales.
- 7. **Customer Gender Revenue Contribution**
  - Based on CustomerContact & Gender column produced in staging.
  - Show Male vs Female revenue distribution.
- 8. **Top 10 Products by Sales Amount**
  - Dynamic Top N using DAX measure.
- 9. **Top 10 Customers**
  - Ranked by total revenue.
- 10. **Category Growth Trend (Quarterly)**
  - Must use time intelligence (Quarter-Year).
  - Highlight growth/decline patterns.

## 5.2 Advanced Analytical Visuals

- 11. **Order Processing Delay Analysis**
  - Visual must show how many days each order took to ship.
  - Required DAX:
    - `ProcessingDays = DATEDIFF(FactSales[OrderDate], FactSales[ShippedDate], DAY)`
- 12. **Discount Impact Analysis**
  - Scatter plot: **Discount vs SalesAmount**
  - Must interpret whether discounts correlate with increased buying.

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

## 13. High-Value Orders Breakdown

- Filter and display orders above a specific threshold (student-computed).
- Example: Orders > 10,000 PKR / USD (depending on freight or total value).
- Must show: Product, Region, Customer, Employee.

## 14. Quantity Heatmap Matrix

- Product (rows) vs Month (columns).
- Show quantity sold patterns across months.

## 15. Employee Productivity by Country

- Must use DimEmployee → DimTerritory → DimRegion.
- Productivity measured by:
  - Total Sales
  - Number of Orders
  - Average Order Value

## 5.3 Mandatory DAX Measures

Students must create the following measures in Power BI:

### Core Sales Measures

- **Total Sales**
- **Total Quantity**
- **Average Sales per Order**

### Customer & Discount Measures

- **Average Discount**

### Operational Measures

- **Processing Days**
- ProcessingDays = DATEDIFF(FactSales[OrderDate], FactSales[ShippedDate], DAY)

### Time Intelligence

- **Year-over-Year Growth**  
(Sales Current Year vs Previous Year)
- **Quarter-over-Quarter Growth (Optional based on data)**

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

# Data Warehouse & Business Intelligence Lab

Instructor: Dr. Khurram Shahzad

BS Data Science Fall 23

- **MTD, QTD, YTD (At least one must be demonstrated)**

## DELIVERABLES (keep it save in your laptop)

Students must submit/have the following:

1. **BPMN Control Flow Diagram** (PDF/PNG)
2. **BPMN Data Flow Diagram** (PDF/PNG)
3. **SSIS Project Folder (Zipped)** containing all packages
4. **Processed SSAS Cube + Screenshots** of tasks (*Only cube screenshots; Power BI uses SQL Server DW*)
5. **Power BI Dashboard (.pbix)**

Good Luck 😊😊😊