E04 – Star Schema (SQL)

Business Intelligence

Exercise

Winter Term 2025/2026

Agenda

- **Introduction**
  - Introduction
  - Notes on software

- **Exercise**
  - Star Schema description
  - Assignments

- **Credits and Materials**

Introduction

- The main goal of the "Star Schema (SQL)" assignment is to give you a practical knowledge on how to use of a typical star schema implementation.

Notes on Software

See the materials from the previous exercise **E02**

Additional instructions on the infrastructure (online platform) for
this course can be found in the PDF
E00 - Infrastructure

Start

- Download supplementary materials "**E04 – Star Schema (SQL) - Supplementary Materials.zip**"

- Check the "data" folder, it should contain SQL file and batch scripts

- You will need a DBMS to work with this data
  - You could use offered infrastructure
    - For further details refer to "**E00 – Infrastructure.pdf**"
  - Note that the data are already available, if not that
    - (option 1) import the data into your database using web UI ("import AdventureWorks into PostgreSQL")
    - (option 2) see the PostgreSQL schema **mtasystem_db_demo**
    - (option 3) import the data use SQL statements or CSV files into your own database

- About data
  - The table fields and data type should be self-descriptive
    - Check the folder with the data as well as the table creation scripts for more details
  - The data is originally designed to be used with PostgreSQL, but you can adapt for other DBMS-es as well

(option 3) Technical Hint : PGAdmin4 and SQL File to Import

It is not always possible to move (e.g. „copy/paste") a lot of data at once between your clipboards and PGAdmin4 SQL tool. Thus, before you will import the data from a big SQL file with inserts, upload this file first to the server. Usually, you should be able to upload files up to 50 Mb.
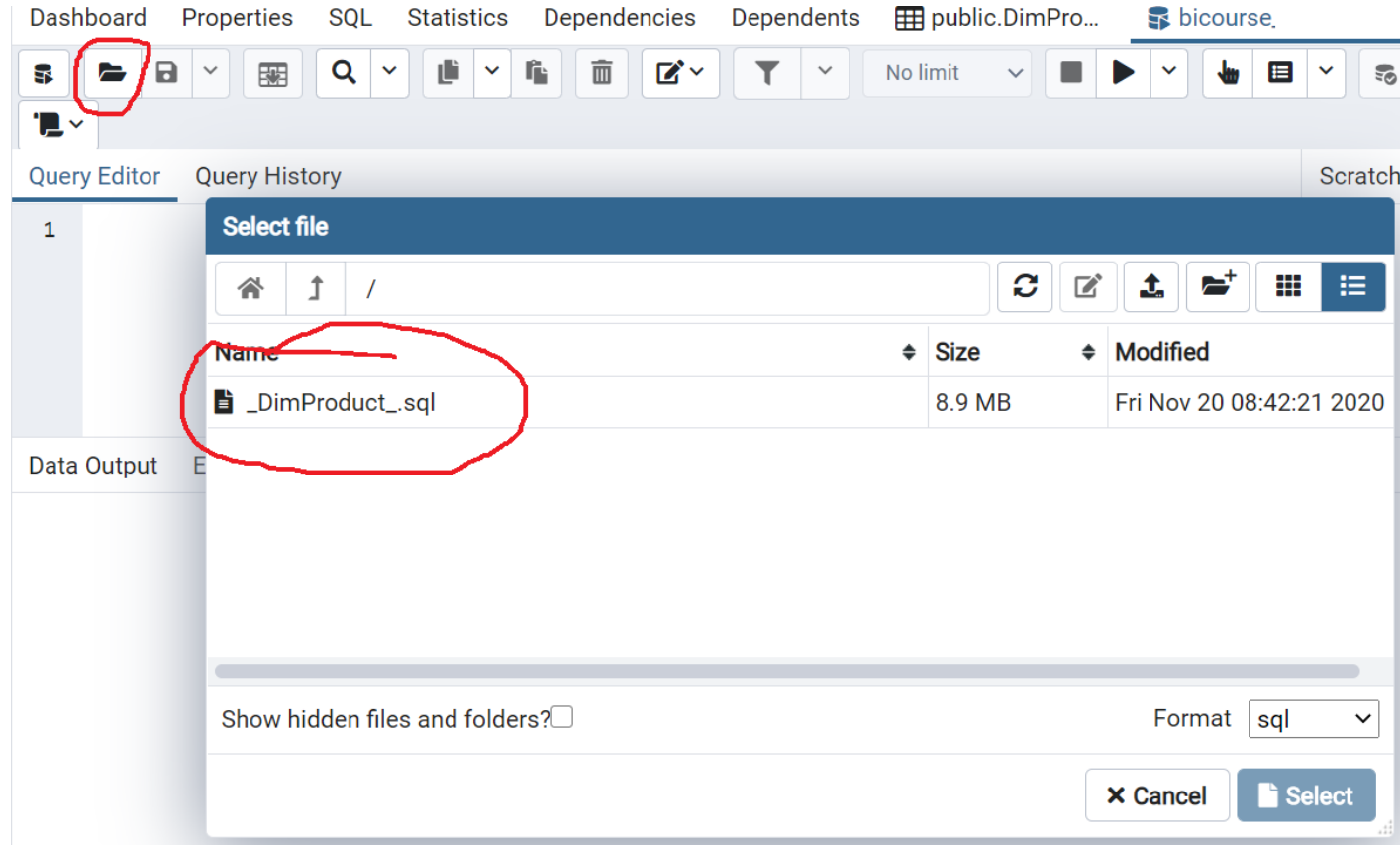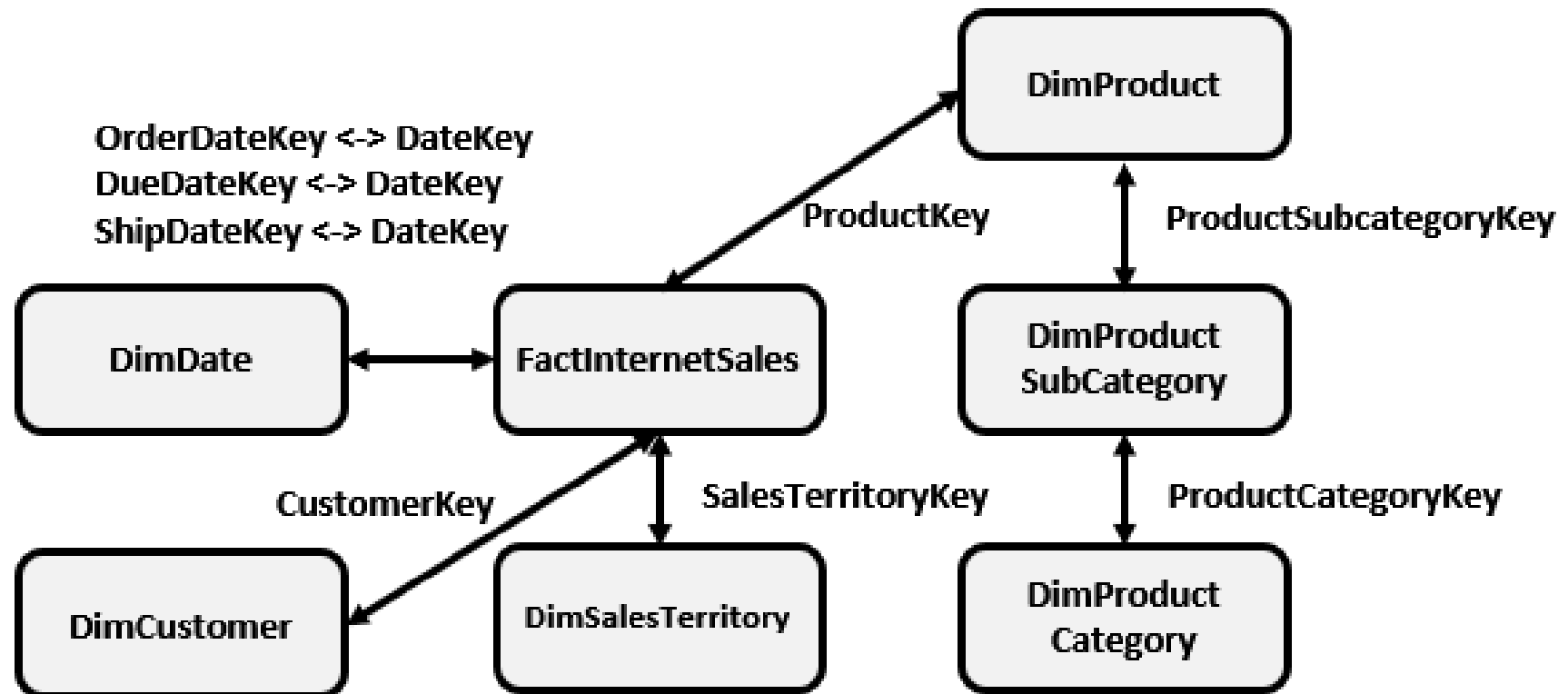
Exercise

Description - Star Schema (Snowflake) of the Internet Sales

AdventureWorks Schema

OrderDateKey <-> DateKey
DueDateKey <-> DateKey
ShipDateKey <-> DateKey

DimProduct

ProductKey

ProductSubcategoryKey

DimDate

FactInternetSales

DimProduct
SubCategory

CustomerKey

SalesTerritoryKey

ProductCategoryKey

DimCustomer

DimSalesTerritory

DimProduct
Category

SQL Query Example That Combines Dimensions

```
SELECT english_product_category_name as Category,
        sc.english_product_subcategory_name as SubCat,
        p.english_product_name as Product,
        SUM(sales_amount) as Sales_Amount
FROM public.fact_internet_sales              s
JOIN  public.dim_product                     p
  on (s.product_key = p.product_key)
JOIN public.dim_product_subcategory sc
  ON (p.product_subcategory_key = sc.product_subcategory_key)
JOIN public.dim_product_category          pc
  ON (sc.product_category_key = pc.product_category_key)
GROUP BY english_product_category_name,
         english_product_subcategory_name,
         english_product_name
```

Tasks

Tasks – Part I

1. Find out how many customers the data warehouse contains?

2. Find out how many products the company has currently?

3. Find out how many facts the fact table contains?

4. What is the most sold product?

5. Which products are the top 3 bestsellers? Bestsellers means "most sold product".

6. On which day of the week the most products are usually sold?

7. On which day of the week the most products are usually shipped?

8. Are there any delays in the process (see due_date and ship_date)?

9. Calculate revenue for each product (use all products). Use ROUND. Sort by revenue and show top 25.

10. Which product make the highest revenue?

Tasks – Part II

1. Who is the best customer (the one that buys at most)? Show customer_key, first_name + last_name, number of sales and revenue.

2. Identify the top 10 customers who could be categorized as "not so good ones" based on their average revenue per order. A "not a good customer" is defined as having a less favorable ratio between the total items bought and revenue. The analysis should only include customers who have placed more than 20 orders. Apply the ROUND function to fields with decimal values in the final results.

3. Compare drop or growth of the sales of each products for the a time period between 01.01.2012 and 31.12.2012. Show top 25 entries, sort by date and select proper columns (product_key, ship_date, due_date, customer_key, unit_price, english_product_name).

4. Are there any customers, who uses multiple locations / stores to buy their products (are enough data provided to answer this question at all)?

5. Are there any product(s), which were sold a lot (e.g., significant > -6) in a particular day/week and then experienced a significant sales drop? Use "ship_date_key" to make sure, that a particular sale has been processed. Hint: you may want to use CTE and LAG to in your SQL query and sort by "product_key", "fiscal_year" and "week_number_of_year".

Tasks – Part III

1. What is the average amount of products / goods / units sold each day over all years? Apply ROUND function to fields with decimal values.

2. What is the average amount of products / goods / units sold each week over all years? Apply ROUND function to fields with decimal values.

3. What is the average amount of products / goods / units sold each month over all years? Apply ROUND function to fields with decimal values.

4. Show sales distribution by months where month name written in French. Order by month_number_of_year and apply ROUND function to fields with decimal values.

Credits and Materials

- *Get started with SQL: Plan and design a database* by Thomas Nield

- Tutorial "Beginner's Guide to Data Modelling" via
http://www.databaseanswers.org/tutorial4_data_modelling/index.htm

- "Industry Data Models" via http://www.databaseanswers.org/data_models/

- MOOC course on databases Introduction to Databases by Jennifer Widom.

- "SQL Exercises, Practice, Solution" via http://www.w3resource.com/sql-exercises/

- "SQLBolt - Learn SQL with simple, interactive exercises" via https://sqlbolt.com/

- SQL tasks with different difficulty levels can be found on HackeRank

Submission

- Use StudIP to upload your solution (PDF report with SQL queries)

- You should upload PDF report with verified SQL queries
  - Further information on the SQL verification system is announced separately

- Name convention for your submission file (without extension)
  - **E04**_FIRSTNAME LASTNAME

- Submission deadline (it is a "soft" deadline)
  - 10 days after this exercise starts
  - Some exercises could take a bit more time and could be submitted later
  - NOTE: to receive feedback, you should first submit your progress