Content

1 Common information

Deadline: submit till 16 January 2025 23.59.

Goal	Required skills	Result
Fully design and develop DB component for an App	Relational DB SQL Data Processing Data Visualization	Schemas, scripts, report, other files

2 Course Work Statement

0. Choose area for development.
1. Design and develop relational database to support functionality of an Application.
2. Design and develop analytical component for an Application.
3. Prepare queries based on OLTP and OLAP DBs to get insights.
4. Document your solution.
5. Submit your solution.

3 Course Work Overview

#	Step	Required result	Mandatory details
0	**Choose area for development**		
	Put chosen topic to Course_Work_Topics_LR_2024.xlsx	Record in xlsx file	One topic per one person
1	**Design and develop all needed DB objects to support functionality of your Application**		
1.1	Develop OLTP solution	Logical schema – picture Tables – SQL script(s)	3NF At least 8 tables
1.2	Prepare data to load to your OLTP database	Several ***.csv** files or one file with several tabs	Data for at least 2 tables No surrogate keys must be present
1.3	Prepare script to load data from CSV to your OLTP database	Script – SQL is preferred	Script should be rerunnable Previously added records should not be overwritten or modified if there are no changes in data

2	**Design and develop data analytical components for your Application**		
2.1	Develop OLAP solution	Logical schema – picture Tables – SQL script(s)	Multidimensional DWH (snowflake) At least 2 Facts At least 1 SCD Type 2 OLAP schema must not duplicate OLTP structure – it should contain some aggregations OLAP DB should be stored separately from OLTP DB
2.2	Develop ETL process to move data from OLTP database to OLAP database	Script – SQL is preferred	Previously added records should not be overwritten or modified if there are no changes in data
2.3	Create visual report based on your OLAP solution	Power BI report	At least one title At least 2 slicers At least 3 visual components used to represent data
3	**Prepare queries based on OLTP and OLAP DBs to get insights**		
3.1	Write queries based on OLTP	Script	At least 2 queries
3.2	Write queries based on OLAP		At least 2 queries
4	**Document your solution**		
	Write down details of implementation	Prepare *.doc file	Overall description of schemas/tables/keys/constraints/relationships Instructions which scripts to run and how to run them (for datasets loading and ETL process)
5	**Submit your solution**		
	Submit all files to GitHub/GitLab and post link to your repository in Moodle	Post link to Moodle **till 16 January 2025 23.59**	▪ Schemas: OLTP, OLAP ▪ OLTP, OLAP scripts ▪ *.csv file(s) with initial data ▪ Script to load data from *.csv to OLTP DB ▪ ETL script to load data from OLTP to OLAP ▪ SQL queries for OLTP and OLAP ▪ Power BI report ▪ *.doc file with description

4 Course Work Specification

#	Step	Result	Description
1.1	Develop OLTP solution – **design 3NF relational DB for full user action flow (8 tables)**	**Schema Scripts**	*ToDo:* Design schema Create tables
			Mandatory details: 3NF At least 8 tables

			Example: ▪ Product search requires categories, subcategories, models, products, brands, manufacturers. ▪ Product view requires product details, product properties, availability status. ▪ With basket user can add and remove items, change items quantity, see items prices and cost, availability status, overall order cost. ▪ Order details display what is being ordered, by whom, where to deliver or pick up, when order is placed and processed, way of payment, order status. ▪ User Account contains user data, login, password, orders, list of liked products. ▪ Admin Actions allow to add new category, product, brand, etc., view and update orders, view and update availability of products.
1.2	Prepare data to load to your OLTP database – **2 datasets**	**File(s)**	*ToDo:* Generate data Save it to *.csv file(s)
			Mandatory details: Data for at least 2 tables No surrogate keys must be present
			Notes: Check the quality, consistency, and format of the data in the CSV file, and make sure that it matches the structure and requirements of the database. Points to consider: ▪ remove any unnecessary or invalid characters, spaces, or quotes; ▪ ensure that the data types, delimiters, and encodings are compatible with the database; ▪ backup the CSV file in case something goes wrong during the import.
1.3	Prepare script to load data from CSV to your OLTP database – **check which data were already uploaded and add only new ones**	**Script**	*ToDo:* Prepare script – SQL is preferred
			Mandatory details: ▪ Script should be rerunnable ▪ Previously added records should not be overwritten or modified if there are no changes in data
2.1	Develop OLAP solution – **design snowflake DWH (2 Facts, 1 SCD Type 2)**	**Schema Scripts**	*ToDo:* Design schema Create tables
			Mandatory details: Multidimensional DWH (snowflake) At least 2 Facts At least 1 SCD Type 2 OLAP schema must not duplicate OLTP structure – it should contain some aggregations OLAP DB should be stored separately from OLTP DB

			Example: Some dimensions could be Customers, Products, and Time. ▪ Dim_Customer may have Customer_ID, Name, Email, Address. ▪ Dim_Product may have ProductID, Name, Category, Price. ▪ Dim_Time may have Date, Month, Quarter, Year. For Type 2 SCD more attributes can be added such as StartDate, EndDate and IsCurrent. Fact_Sales could include: ▪ Quantity_sold (a measure) ▪ Total_sales (a measure) ▪ ProductID (a foreign key related to the Dim_Product) ▪ CustomerID (a foreign key related to the Dim_Customer) ▪ Date (a foreign key related to the Dim_Time)
2.2	Develop ETL process to move data from OLTP database to OLAP database – **check which OLTP data were already uploaded and add only new ones, made transformations if needed, save data to DWH**	**Script**	*ToDo:* Prepare script – SQL is preferred
			Mandatory details: Previously added records should not be overwritten or modified if there are no changes in data
			Notes: 1. Identify reference OLTP data: write a query/few queries that defines the set of permissible values your DWH may contain. For example, in a country data field, specify the list of country codes allowed. 2. Extract data from the source: convert it into a single format for standardized processing. 3. Validate data: keep data that have values in the expected ranges and reject any that do not. For example, if you only want dates from the last year, reject any values older than 12 months. 4. Transform data: remove duplicate data (cleaning), apply business rules, check data integrity (ensure that data has not been corrupted or lost), and create aggregates as necessary. For example, if you want to analyze revenue, you can summarize the dollar amount of invoices into a daily or monthly total. You may need to program numerous functions to transform the data automatically. 5. Stage data (optional): sometimes it is better not to load transformed data directly into the target data warehouse. Instead, data first enters a staging database which makes it easier to roll back if something goes wrong. 6. Publish data to your data warehouse: load data to the target tables.
2.3	Create visual report based on your OLAP solution – **create meaningful Power BI report answering analytical questions**	**PBI Report**	*ToDo:* Connect Power BI to your DWH Download data Prepare your data with a few transformations Create Power BI report
			Mandatory details: At least one title At least 2 slicers At least 3 visual components used to represent data

	regarding your topic		*Example:* Data transformation: ▪ change data types: decimal to whole number ▪ change data view: from lowercase to uppercase ▪ filter data Visual components: ▪ create a line chart to see which month and year had the highest profit ▪ create a map to see which country/region had the highest profits ▪ create a bar chart to determine which companies and segments to invest in ▪ create two different slicers to narrow in on performance for each month and year
3	Prepare queries based on OLTP and OLAP DBs to get insights	**Script**	*ToDo:* Prepare script with queries
			Mandatory details: At least 2 queries for OLTP and 2 for OLAP
			Example: Ask a question about your data – for example for Bike Rental App – how many and which bikes are rented by weeks/month in years. Answer it by using OLTP tables and then OLAP tables.