Content

1 Common information

Deadline: submit till 9 June 2024 23.59.

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

2 Course Project Statement:

0. Choose area for development.
1. Design and develop all needed DB objects to support functionality of an Application.
2. Design and develop analytical component for an Application.
3. Document your solution.
4. Submit your solution.

3 Course Project Overview

#	Step	Requirement	Mandatory details
0	**Choose area for development**		
	Put chosen topic to DB_course_project_topics.xls file	E-commerce App or Rental App + sphere	One topic per one person
1	**Design and develop all needed DB objects to support functionality of an Application**		
1.1	Design ER-diagram	Show main entities and relationships between them	
1.2	Develop OLTP solution	Present logical schema, physical schema, tables, indexes, queries for main screens, functions and procedures for main actions, triggers (if needed), roles and rights	3NF At least 8 tables
1.3	Prepare data to load to your OLTP database	Create *.csv files or one file with several tabs	At least 2 datasets

1.4	Prepare script to load data from CSV to your OLTP database	Script should be rerunnable	Previously added data should not be rewritten
2	**Design and develop analytical component for an Application**		
2.1	Develop OLAP solution	Present logical schema, physical schema, tables	Multidimensional DWH At least 2 Facts At least 1 SCD (Type 2)
2.2	Develop ETL process to move data from OLTP database to OLAP database	Script for ETL should be rerunnable	Previously added data should not be rewritten
2.3	Create visual report based on your OLAP solution	Prepare visual report, report template or dashboard	At least one title At least 2 filters At least 3 visual components
3	**Document your solution**		
	Write down details of implementation	Prepare *.doc file	Overall description of schemas Overall description of main functions, procedures Instructions which scripts to run for datasets loading and ETL process and how to run them
4	**Submit your solution**		
	Submit all files to Git/cloud and post link to it in Moodle	Post link to Moodle **till 9 June 2024 23.59**	▪ Schemas ▪ OLTP scripts ▪ *.csv file(s) with initial data ▪ Script to load data from *.csv to OLTP DB ▪ OLAP scripts ▪ ETL script ▪ Power BI report ▪ *.doc file with description

4 Course Project Specification

#	Step	Result	Description
1.1	Design ER-diagram	**Diagram**	*Theory:* An Entity Relationship (ER) Diagram is a type of flowchart that illustrates how "entities" such as people, objects or concepts relate to each other within a system. ERDs or ER Models use a defined set of symbols such as rectangles, diamonds, ovals and connecting lines to depict the interconnectedness of entities, relationships and their attributes. They mirror grammatical structure, with entities as nouns and relationships as verbs.
1.2	Develop OLTP solution – **design 3NF relational DB for full user action flow (8 tables)**	**Schemas Scripts**	*Theory:* An OLTP database stores and manages data related to everyday operations within a system or a company. OLTP typically deals with query processing (inserting, updating, deleting data in a database), and maintaining data integrity and effectiveness when dealing with numerous transactions simultaneously. OLTP queries are simple and typically involve just one or a few database records.
			ToDo: Design schemas Create needed tables, indexes Save queries for main screens Create functions and procedures for main actions Add roles to distinguish user and admin
			Example: Product search (categories, subcategories, models, products, brands, manufacturers) Product view (product details, product properties, availability status) Basket (user can add and remove items, change items quantity, see items prices and cost, availability status, overall order cost) Order details (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 (user data, login, password, orders, list of liked products) Admin Actions (add new category, product, brand, etc., view and update orders, view and update availability of products)
1.3	Prepare data to load to your OLTP database – **2 datasets**	**File(s)**	*Theory:* CSV file is a common format for storing and exchanging tabular data. A CSV file contains rows of values separated by commas, and each row represents a record or a tuple in a table.
			ToDo: Generate 2 datasets Save them to *.csv file(s)
			Example:

			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.
1.4	Prepare script to load data from CSV to your OLTP database – **check which data were already uploaded and add only new ones**	**Script**	*ToDo:* You can use built-in RDBMS commands, graphical user interfaces, or external tools for development of a data load pipeline.
			Example: ▪ 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.
2.1	Develop OLAP solution – **design snowflake DWH (2 Facts, 1 SCD Type 2)**	**Schemas Scripts**	*Theory:* OLAP business intelligence queries often aid in trends analysis, financial reporting, sales forecasting, budgeting and other planning purposes. A Dimension table is a structure that categorizes facts and measures in order to enable users to answer business questions. Dimensions are descriptive and define the characteristics of a business object. They provide context to facts – as they hold the fields which are descriptive, qualitative and textual. SCD - Slowly Changing Dimensions is a concept that is used to enable the historic aspect of data in an analytical system. SCD of Type 2 means history will be added as a new row. A Fact table contains the quantifiable data for analysis — the numerical measures (often additive) of the business processes. The Fact table also has foreign keys which refer to candidate keys in the Dimension tables.
			ToDo: Design schemas Create needed tables, indexes Save queries for future visual report
			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**	*Theory:* ETL, which stands for extract, transform, and load, is the process data engineers use to extract data from different sources, transform the data into a usable and trusted resource, and load that data into the systems end-users can access and use downstream to solve business problems.
			ToDo: You can use any programming language or software solution for development of an ETL pipeline.
			Example: 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 regarding your topic**	**Report**	*Theory:* Visualizations display insights that are discovered in the data. A visual report might have a single page with one visual or it might have pages full of visuals.
			ToDo: Connect Power BI to your DWH Download data Prepare your data with a few transformations Build a report with a title, two filters and three visuals
			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