DBAS 1007 Project – Data Migration

# Project Introduction

In this database project, you will take an existing, populated database, re-design parts of it according to the provided requirements, implement those design changes, and then migrate all existing data from the original database to the newly-enhanced database. Part of the data migration will entail data manipulation to conform to the new design. Finally, you will create some test scripts to verify the accuracy of your data migration.

This project is intended to evaluate you on all the core DBAS course concepts, so will require analysis and design of an ERD, implementation of both Data Definition Language and Data Manipulation Language scripts, as well as using queries in a variety of ways to help transform, populate and test the validity of your work.

# Project Evaluation & Deadline

Due Date: **See due date designated on the Final Project dropbox on Brightspace.**

This project is worth 20% of your overall course mark. There are three deliverables:

1. An ERD design containing the required changes/enhancements to the Northwind database.
2. A **SINGLE script** containing DDL and DML statements to create all database objects for the enhanced Northwind database and migrate and/or populate all tables with the existing data from the original Northwind database.
3. A second Test Plan script, containing queries used to verify the accuracy of data migration.

|  |  |
| --- | --- |
| Deliverable | Value |
| ERD Document | **25%** |
| DDL & DML script | **50%** |
| Test Plan script | **25%** |

# Required Materials

To complete this project, you will need a copy of the Northwind ERD and database creation scripts, which are available on Brightspace. The software programs required for this project are Visual Paradigm, MariaDB, and DataGrip.

# Marking & Submission

File deliverables will be uploaded to the Final Project dropbox on Brightspace. Since the submission includes multiple files, please put all files into a single ZIP file, named similar to: **DoeJohn\_DBASProject.zip**

# Project Scenario & Requirements

# A fictional company named *NorthWind Traders* has hired you to make some enhancements to their existing order-tracking database. You will be required to analyze their existing database, design a solution to meet their requirements, create the new database according to the new design, and then migrate all data from their existing database into the new one. They would also like some basic query scripts to verify the success of the data migration.

### **Enhancement #1 – Better Inventory Tracking**

The existing database has three fields in the Products table that are currently used to track inventory stock (UnitsInStock, UnitsOnOrder, ReorderLevel). However, the company owns three different warehouses, and currently has no way of tracking stock levels for each product in each warehouse. Inventory for any product could be stored at any of the three warehouse locations. Each warehouse has a name and a street address that should be stored in the new database. The warehouses are commonly known by their location within the city: The Dockside warehouse, the Airport warehouse and the Central warehouse. The new scheme should be able to record the same basic data as the existing database (UnitsInStock, UnitsOnOrder, ReorderLevel), but also reference the warehouse in which the products are currently being stored in, or re-ordered from. **Note:** Your script should distribute existing products amongst the three warehouses. Feel free to make up the distribution.

**Enhancement #2 – Centralization of Country Data**

The accounting team often uses country-related data from the database when generating reports based on financial, supply chain or human resources data tracked in the system. The current database records this country data in several tables (Customers, Orders, Suppliers, & Employees). However, frequent report errors are being encountered due to problems with the country data. For example, if a country name is misspelled, it shows up as a separate report item, even though it is obviously the same country with a typo in the name.

The accounting team has proposed a centralized list of countries in the database. Each of the tables that currently record Country data should no longer allow user-entered country names, but should instead “look up” the desired country from the new central Country table. During data entry via the website, the country will be selected from a dropdown list instead of being typed manually. The software dev team will implement that change on the website, but the new database design should support a centralized Country table.

**Enhancement #3 – Full Names with Title**

The HR department has requested a database change to help integration with the new HR software program. The new software does not work well when importing separate values for first/last names and the title of courtesy. They propose adding a new field that contains a combined value in the following format: <Courtesy Title> <First Name> <Last Name>. Examples: Mr. John Doe or Dr. Jane Smith. The proposed field will not replace the usage of the existing separate fields; it will simply be a redundant field containing the full value from the three other fields, for use only with the new HR software program.

# Deliverable Descriptions & Requirements

Database Setup & Requirements

For this project, you will need to create two new databases:

1. From Brightspace, download and run the script **Northwind\_CreateOrigDB.sql** to create the client’s “original” database, which will be created in MariaDB as **“northwind\_orig”.** This db represents the “untouchable” database you received from your client.
2. Create a new, empty database in MariaDB named **exactly**: **northwind\_Enhanced.** This dbrepresents the “new and improved” db you will give your client upon completion of all work and testing.

After the initial database creation, no structural or data changes (DDL or DML) should occur in the original Northwind database.

Of course, you can still use read-only statements (such as SELECT) against the original database as part of your development, and in the DML and Testing portions of the project.

## Entity Relational Diagram (ERD)

For your ERD submission, please take the provided Northwind ERD and alter it in Visual Paradigm to include your changes.

The ERD submitted for this deliverable should meet all the following requirements:

* **Submit your completed ERD as an image or .pdf file.**
* **Highlight all changes from the original db using either a red circle or red text.**
* Include all entities and attributes needed from the original database
* Include all appropriate relationships (including recursive or hierarchical, if applicable).
* Include a primary key for every entity.
* Include a foreign key for every relationship.
* Matching datatypes and sizes for all foreign keys with their associated primary keys.
* Appropriate data-types and sizing for all fields.
* Appropriate decisions on which fields should be mandatory or not.
* Crow’s-foot notation and physical data-types displayed.
* Where applicable, other constraints may be indicated in your ERD.

## Data Definition Language (DDL)

The DDL portion of the script submitted for this deliverable should meet all the following requirements:

1. You may use the generated DDL script from the original database as a base for your own work, however, the standard method for work of this type would include ALTER statements to enact any structural changes to the database.
2. An ERD is a blueprint for a database. The data structures created by your scripts should match your ERD **exactly** in every manner, including structure, relationships, table/field names, datatypes and constraints.
3. Create or alter all entities, attributes, constraints and datatypes present in your new ERD.
4. Create or alter all objects in an order that the full script can be run as a whole, with no errors.
5. Create or alter all key constraints (Primary and Foreign Keys), as identified in your ERD.
6. Create or alter all appropriate data integrity constraints (NOT NULL, CHECK, DEFAULT, UNIQUE.). There should be at least one example of each data integrity constraint type.
7. Create DROP statements for all applicable database objects.
8. If temporary columns are added for use during the data migration, they should be removed later in the scripts.

## Data Manipulation Language (DML)

The DML portion of the script submitted for this deliverable should meet all the following requirements:

1. When populating the enhanced database with records, **you may NOT just copy the INSERT statements from the provided original script**. That is not a realistic method, since the database would be in use during the design/development of this fictional project. If you add a new table, you may use plain INSERT statements to populate it, but for existing tables, you must use statements such as **INSERT INTO (SELECT…), UPDATE** etc. to make any data additions or changes required.
2. **All** data from the original database should be properly inserted into your newly designed database.
3. In some cases, new data that did not exist in the previous database may need to be added.
4. There may be some data in the original database that does not meet the constraints in your new database. In these cases, the data will need to be manipulated or transformed during insertion.
5. All data should be added in a single running of your DML script, so the order in which tables are populated should be appropriately considered.
6. All data used for primary key fields should be unique between all records in a particular table, and should not contain any missing values.
7. All data used for foreign key fields should demonstrate the relationship to related data in the associated parent table, and match in datatype and size.
8. Where applicable, DML scripts should show proper, dynamic of use sequences.
9. All data records added to the database should adhere to all data integrity constraints established in the design.

## Test Plan Script

The Test Plan script submitted for this deliverable should meet all the following requirements:

1. Each test case should be commented in the script, to indicate the **purpose** for the test, and the **expected results**.
2. Include queries to compare and verify record counts between ALL tables in the original and new databases.
3. Include queries to verify there are no excluded records between ALL tables in the original and new databases.
4. Include queries to ensure that all new tables added to the new database have data in them

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DBAS 1007 Project - Data Migration - ERD** | | | Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| ***Level *** | ***Insufficient*** | ***Developing*** | ***Sufficient*** | ***Excellent*** | ***Points Earned*** |
| ***Element* ** | ***0 points*** | ***1-2 points*** | ***3-4 points*** | ***5 points*** |
| ***Entities*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Sufficient entities added to ERD to meet all required design changes. New entities integrate with old design seamlessly. |  |
| ***Attributes*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Sufficient attributes added to new and existing entities to meet all required design changes. All attribute names, datatypes and sizes done properly. |  |
| ***Primary Keys*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All new and existing relationships are correctly defined using proper primary keys, with appropriate datatypes. |  |
| ***Foreign Keys*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All new and existing relationships are correctly implemented with proper foreign keys. |  |
| ***Professionalism*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Completed ERD displays all of the following: -All changes highlighted red - A professional look  - A logical & readable layout - Datatypes displayed - Uses Crow Foot notation. - Properly labelled. |  |
|  |  |  | **Total Points:** |  | **/25** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DBAS 1007 Project - Data Migration - DDL & DML** | | | Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| ***Level *** | ***Insufficient*** | ***Developing*** | ***Sufficient*** | ***Excellent*** | ***Points Earned*** |
| ***Element* ** | ***0 points*** | ***1-2 points*** | ***3-4 points*** | ***5 points*** |
| ***Table Creation*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All entities defined in the ERD are properly created. Script includes appropriate DROP statements |  |
| ***Enhancements*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All tables and fields required for the db changes are properly created. |  |
| ***Key Constraints (DDL)*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All tables have a primary key constraint correctly applied to the key field. All defined foreign key relationships are enforced using properly-applied FK constraints. |  |
| ***ALTER Statements*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | ALTER statements used in a correct and appropriate fashion wherever required. |  |
| ***Existing Data*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All pre-existing data from the original database is present and correct. |  |
| ***New Data*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All new data required to be added for the changes present and correct. |  |
| ***Foreign Key & Data Relations (DML)*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Foreign key constraints followed, and all data correctly establishes links between related records, both existing and new. |  |
| ***Data Transformations*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | All required data transformations or manipulations correctly done. |  |
| ***Script Usability*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Overall script constructed logically, in the proper order and is able to be run without error. |  |
| ***Script Commenting*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Script is extensively commented, to both explain and organize statements. |  |
|  |  |  | **Total Points:** |  | **/50** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DBAS 1007 Project - Data Migration - Test Plans** | | | Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | |  |
| ***Level *** | ***Insufficient*** | ***Developing*** | ***Sufficient*** | ***Excellent*** | ***Points Earned*** |
| ***Element* ** | ***0 points*** | ***1-2 points*** | ***3-4 points*** | ***5 points*** |
| ***Record Count Tests*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Statements included that thoroughly and correctly compare and count records for all tables between the original and new databases. |  |
| ***Record Exclusion Tests*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Statements included that thoroughly and correctly test for any missing or excluded records for all tables between the original and new databases. |  |
| ***New Table Data Tests*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. | A good effort, but submission has a few issues and/or areas for improvement. | Statements included that thoroughly and correctly test that all newly-created tables contain the correct data. |  |
| ***Commenting (x2)*** | Missing entirely, or contains too many issues and/or omissions. | An effort was made, but submission has multiple issues, major issues, or omissions. (2-4 pts) | A good effort, but submission has a few issues and/or areas for improvement. (6-8 pts) | Script and test plans are extensively commented, to explain purpose and expected results of each test. (10 pts) |  |
|  |  |  | **Total Points:** |  | **/25** |