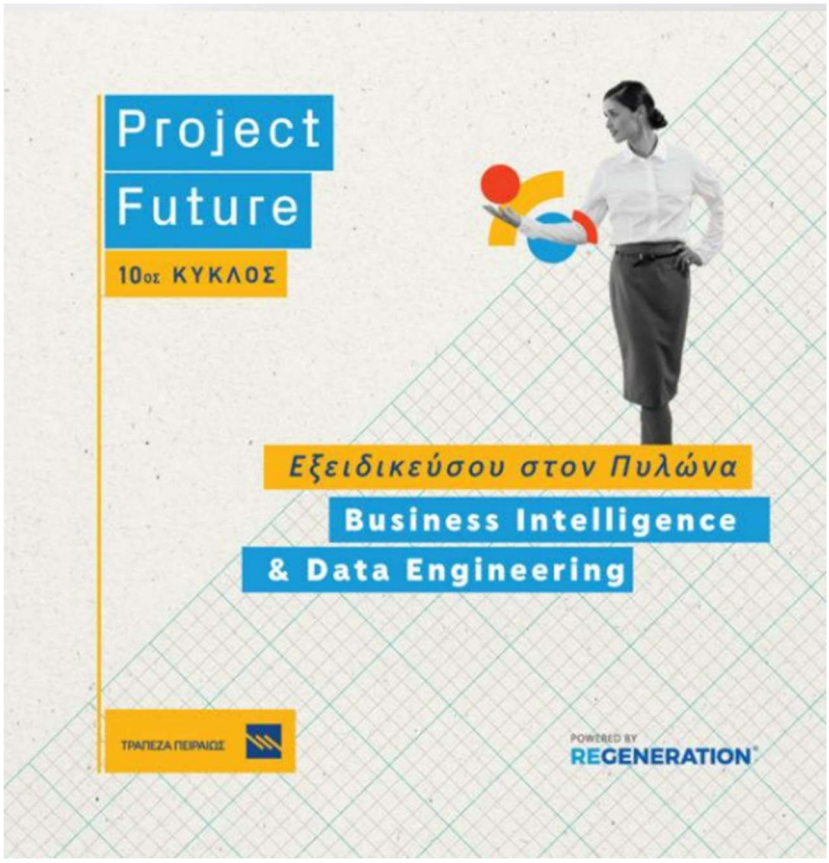

Project assignment
Business Intelligence and Data Engineering

Regeneration Academy

Trainer Partner

Code.Hub

Project Future 10ος ΚΥΚΛΟΣ

REGENERATION®

Project Assignment: Melodica Business Intelligence Pipeline

Introduction

Melodica Media Corp. is a multinational company headquartered in California, US that operates the namesake *Melodica* online music store. The store is comprised of downloadable MP4 files of songs that customers can download a piece.

The company has contracted you and your team to create a **business intelligence pipeline** to analyze **track sales** and support the company's decision-making process.

Project Scope

You are expected to create a complete BI pipeline, consisting of a DW, SSIS packages to automate the creation and loading of the DW as well as a report which will load data directly from a Tabular model based on your DW.

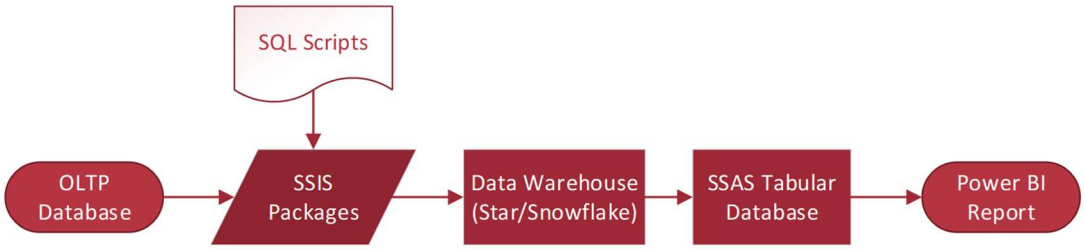

Figure 1. Flowchart representation of the expected BI pipeline.

Deliverables

Your pipeline should include all the following components:

A1.) A Data Warehouse (Star/Snowflake Schema) based on the data found in the provided OLTP database, using an intermediate staging area that supports:

A2.) SCD Type 2 for the table containing customers, supporting retention of historical values for customer data.

A3.) Incremental loading (delta loading) of new rows for the fact table.

You are expected to deliver:

- SQL Scripts to create and incrementally load the Staging Area
- SQL scripts to create the DW and load it from the staging area

- A *.bak* file containing a backup of the loaded DW.

B.) An SSIS project with packages to:

B1.) Create the DW and Staging Area.

B2.) Perform incremental loading of the staging area and load new data into the DW, as well as perform any maintenance you deem necessary.

You are expected to deliver:

- A Visual Studio solution containing the described SSIS project.

C.) An SSAS Tabular model based on the DW you created. The model should contain all data from the DW as well as any DAX measures or calculated columns you consider useful.

You are expected to deliver:

- A Visual Studio solution containing the described SSAS Tabular project.
- A *.abf* file containing a backup of the finalized Tabular database.

D.) A Power BI report to analyze sales, which will draw data through a live connection with the tabular model you created in the previous step.

You are expected to deliver:

- A *.pbix* file containing your work.

E.) Executive summary of no more than a page including the main and important topics of your work.

F.) A PowerPoint presentation summarizing your work, with bullets describing the purpose of your project, any assumptions, decisions you made as well as their justifications, and points of future improvement or expansion. This presentation will be used during the final presentation of your project, by the **whole team**.

All files can be delivered directly though *Microsoft Teams* in your team's *File* section or by providing links to an accessible GitHub repository or a text file also uploaded on your team's *Files* section.

Data Description

You are provided with a complete OLTP database, *Chinook* which will provide the data for your project. For the purposes of this scenario, you will consider the *Chinook* database as the OLTP database maintained by *Melodica* and its digital storefront.

The database contains data about the songs sold by the company. Each song *(Track)* has an *Album* and each album has an *Artist*. Tracks also have a genre and may appear in *Playlists*. A sale is represented by an *Invoice* and each invoice can include multiples of the same track purchased, as well as many different tracks. Each *Invoice* is produced for a *Customer,* who is also assigned an *Employee* as a support representative.

For your convenience and support, you are also provided with a schema of the OLTP database in the last appendix of this handout.

Evaluation Criteria

Your project will be evaluated based on three pillars:

- **Your understanding of the business problem and ability to provide solutions** (correct schemas, useful measures or calculated columns, useful, organized, and readable reports).
- **Technical accuracy and performance** (correct use of tools, absence of bugs, use of best practices for performance and stability).
- **Project management.** How the team organizes the work and divides it into tasks. An online pm tool is recommended.

Appendix

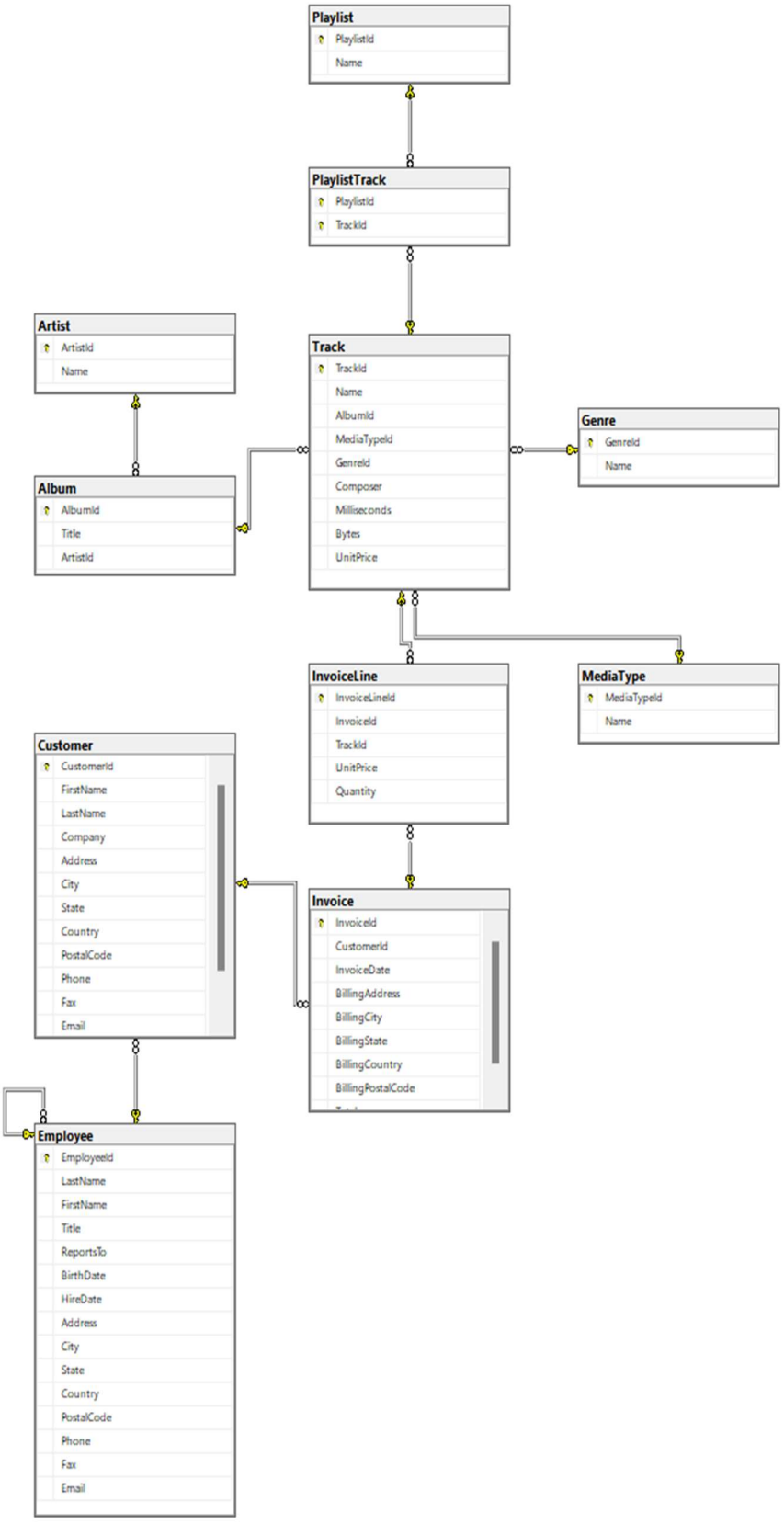

Figure 2. Diagram of the Chinook OLTP database schema.