# iMedia Data Pipeline Design

## Instructions:

1. Design the data pipeline for the music database.  
2. Follow the structure provided to detail each phase of the data pipeline.  
3. Save the document as 'iMedia\_Data\_Pipeline\_Design\_Completed.docx.'

## Automate:

1. Very Script Started
2. Clear existing tables before each run
3. Validate Successful/Unsuccessful Script Run
4. Verify Script Ended
5. Save Script
   1. Make Script Executable
   2. Run Script

## Extract Phase:

1. Data Sources and Extraction Methods:

\* Artists: CSV file

\* Albums: CSV file

\* Tracks: CSV file

\* Genres: CSV file

## Transform Phase:

1. Data Cleaning Steps:

\* Step 1: Handling missing values in artist data

\* Step 2: Removing duplicate entries in album data

\* Step 3: Correcting data formats in track metadata

2. Data Transformation Steps:

\* Step 1: Converting birthdates to a standard format

\* Step 2: Aggregating track play counts by album

\* Step 3: Mapping genres to standardized genre IDs

## Load Phase:

1. Target Database: SQLite database

2. Loading Strategy:

\* Insert cleaned and transformed data into the target tables within the database:

- Artists: Map ArtistID, Name, BirthDate, and GenreID to the Artists table.

- Albums: Map AlbumID, Title, ReleaseDate, and ArtistID to the Albums table.

- Tracks: Map TrackID, Title, Duration, and AlbumID to the Tracks table.

- Genres: Map GenreID and Name to the Genres table.