# 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\_Template.docx.'

## Automate:

\* [Describe the steps to automate the ETL script, e.g., Initialization, Clearing Existing Data, Validation, etc.]

## Extract Phase:

1. Data Sources and Extraction Methods:

\* Artists: [Describe the data source and extraction method, e.g., CSV file]

\* Albums: [Describe the data source and extraction method, e.g., API]

\* Tracks: [Describe the data source and extraction method, e.g., Excel file]

\* Genres: [Describe the data source and extraction method, e.g., Text file]

## Transform Phase:

1. Data Cleaning Steps:

\* Step 1: [Describe the first data cleaning step, e.g., handling missing values in artist data]

\* Step 2: [Describe the second data cleaning step, e.g., removing duplicate entries in album data]

\* Step 3: [Describe the third data cleaning step, e.g., correcting data formats in track metadata]

2. Data Transformation Steps:

\* Step 1: [Describe the first transformation step, e.g., converting birthdates to a standard format]

\* Step 2: [Describe the second transformation step, e.g., aggregating track play counts by album]

\* Step 3: [Describe the third transformation step, e.g., mapping genres to standardized genre IDs]

## Load Phase:

1. Target Database: [Specify the target database, e.g., SQLite database]

2. Loading Strategy: [Describe the loading strategy, e.g., batch loading, real-time loading]