**Abuchi Okeke**

**Big Data (Enhance IT)**

**10/04/2020**

Domain: **Music**

Spotify is used for music streaming and it contains over 50 million songs. The Spotify Web APIs can also be used for data extraction and analysis purposes. This project will integrate Spotify Web APIs for data extraction based on different playlists using Spotipy (python library). The datasets will be stored in a .csv files and will be ingested into different tables in three separate database servers (MySQL, PostgreSQL, SQL Server) as illustrated in figure 1.0 below. The end goal is to generate large scale music datasets that can be used for Big data analysis.

Project Design

Database

**MySQL DB Server**

**PostgreSQL**

Tables

- Genre

- Subgenre

- Year

Tables

- PlayList

- PlayListTrack

- Tracks

Tables

- Album

- Artist

- Edges

**SQL Server**

- Open source db/datasets

- Random data generation

Spotify API integration

(Library: python spotipy)

Datasets Extraction

Web API

(Spotify)

Figure 1.0: Project process illustration

**Tools**

* DBeaver
* Python Libraries
* [Google Colab](https://colab.research.google.com/)/Pycharm/Jupyter Notebook

Click here for python Spotify Web API integration:

<https://github.com/Buchiexplores/Big_Data_Training/blob/master/Assignment5/Scripts/Spotipy_Data_Extraction.ipynb>

Click here for the datasets:

<https://github.com/Buchiexplores/Big_Data_Training/tree/master/Assignment5/Datasets>

Click here database servers’ scripts:

<https://github.com/Buchiexplores/Big_Data_Training/tree/master/Assignment5/Scripts>