

Team 1 - Spotify Music Management System

Gracy Ahuja, Fahad Ibrahim, Sanjana Kumar, Prachi Shah

Design Analysis-

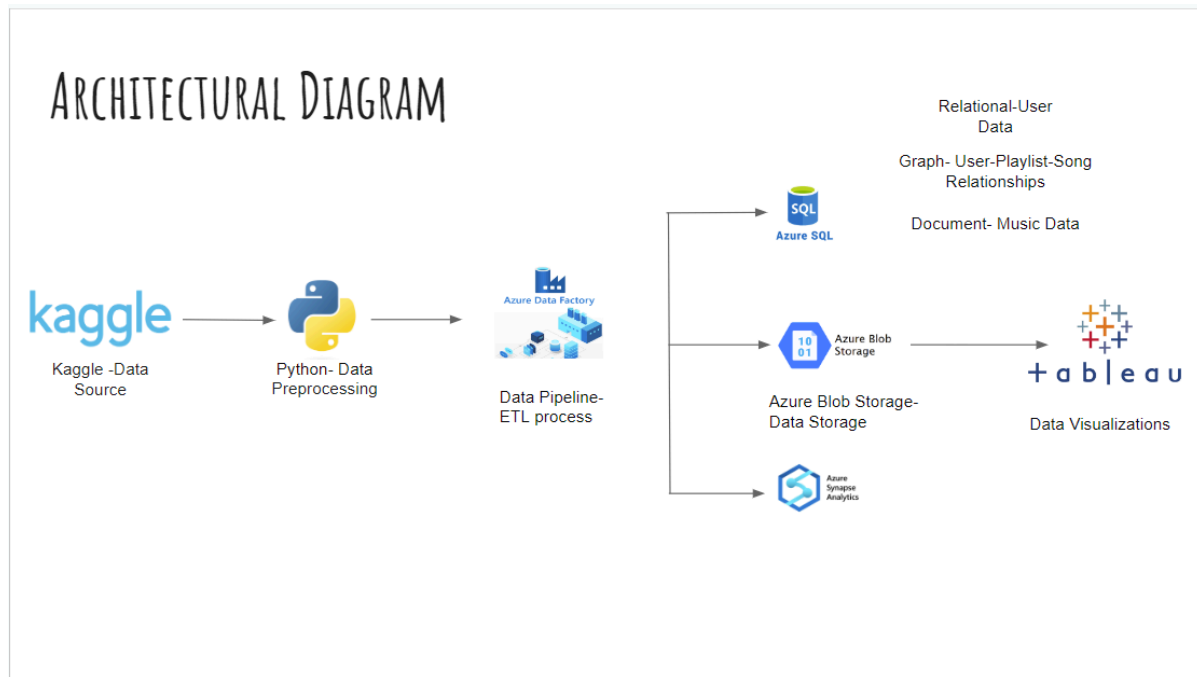
We will explore datasets on Kaggle and focus on those containing user behavior and music related information. For data we can not find, we are planning to create our own using Mockaroo. We will utilize Pandas in Python to preprocess and combine it with the existing datasets from Kaggle.

For the ETL processes, we plan to use the Azure Data Factory(ADF), as ADF will allow us to move the databases from different files to one destination of our choice. Also using ADF we will be able to apply various transformations on our datasets such as filtering songs as per their artists, sorting the music as per playlists, etc. Apart from the ETL processes, ADF will also help in seamless integration with other Azure Services.

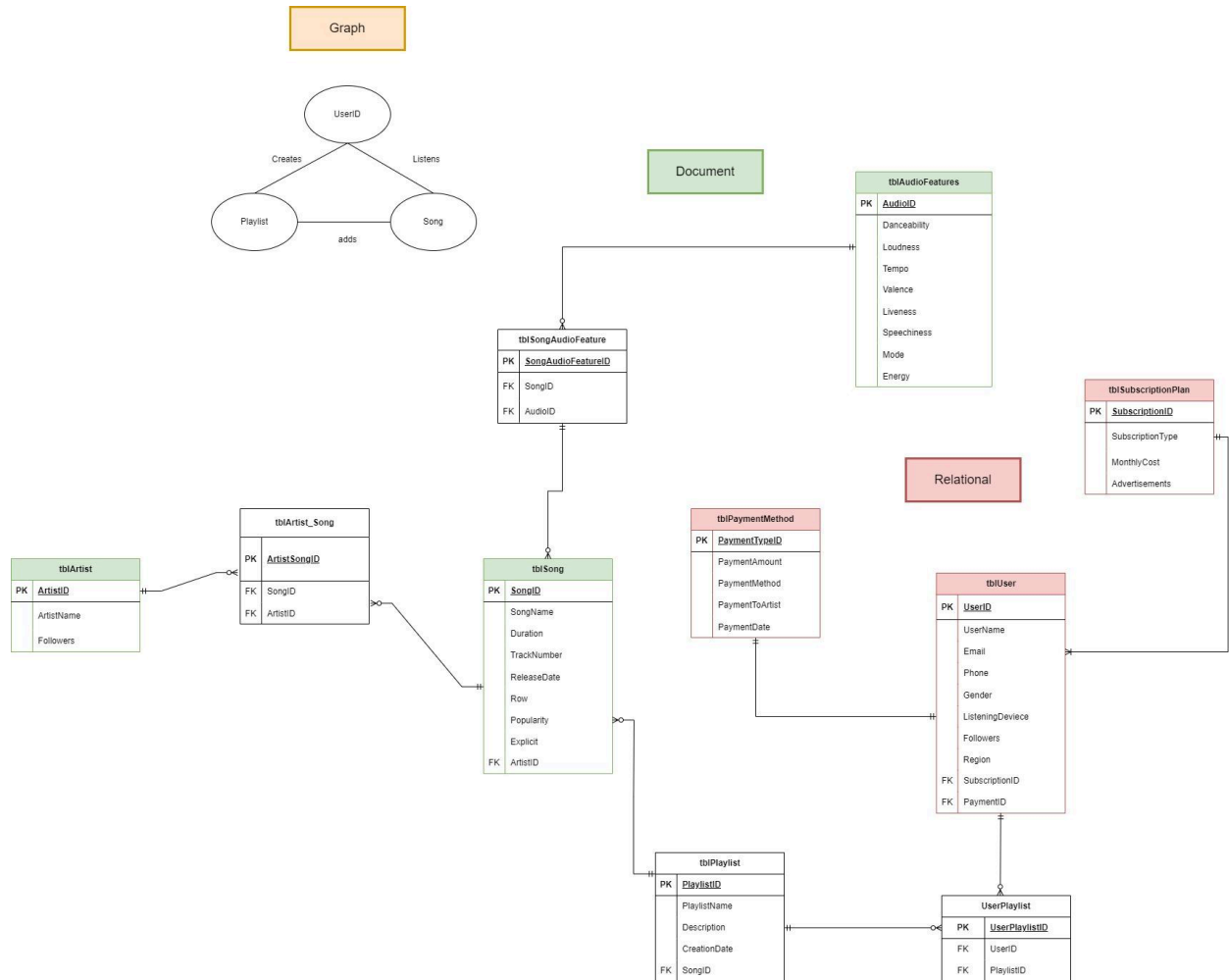
After we have cleaned the files and after the ETL data pipeline process is completed, The data is stored in Azure SQL, Azure Blob Storage, and Azure Synapse. Once the data has been stored and completed, we can do further processing and analysis to extract insights. Additionally, data quality and validation procedures to ensure the accuracy and reliability of the data.

The data models we will be using are Graph, Document and Relational. The data stored in the graph data model is user-playlist-song relationships. The document database mainly focuses on storing the music data and relational databases will store the transactional data of the user such as subscriptions and payments. After the database is created, we are planning to create visualizations in Tableau and get insights on the Spotify Music Management System.

Architecture Diagram-



Document, Graph, and Relational Model-



SPOTIFY MUSIC MANAGEMENT SYSTEM

ERD + ARCHITECTURAL DIAGRAM



Presented By

Fahad Ibrahim, Gracy Ahuja, Sanjana Kumar, Prachi Shah

OBJECTIVES

- **Retrieve data:** Gather from various sources for insights into user behavior and content popularity.
- **Store data:** Organize efficiently for easy retrieval and analysis.
- **Visualize data:** Create interactive graphs for easy understanding of trends and engagement patterns.
- **Identify influencers:** Analyze user interactions to find key figures and trending artists.
- **Track engagement:** Monitor user interactions to improve user experience.
- **Analyze trends:** Identify popular genres and emerging artists for curation.
- **Geographic insights:** Understand regional preferences for targeted recommendations.
- **Manage data volumes:** Ensure smooth streaming experiences globally.

ARCHITECTURAL DIAGRAM

kaggle

Kaggle -Data
Source



Python- Data
Preprocessing



Data Pipeline-
ETL process



Relational-User
Data

Graph- User-Playlist-Song
Relationships

Document- Music Data



Azure Blob Storage-
Data Storage



Azure
Synapse
Analytics



Data Visualizations

ENTITY DESCRIPTIONS

1. **tblUser**- Represents individual Spotify users and stores their information.
2. **tblPlaylist**- Contains lists of songs created by users.
3. **tblAlbum**- Stores information about music albums available on Spotify.
4. **tblSong**- Holds data about individual songs, including title, duration, and artist information.
5. **tblGenre**- Categorizes songs based on music style or genres.
6. **tblArtist**- Stores details about music artists.
7. **tblAudioFeature** - Describes the features of a particular audio in terms of loudness, valence, danceability, etc.
8. **tblSubscriptionPlan** - Represents different types of plans. Eg- Premium, Family, etc
9. **tblPaymentMethod** - Describes the mode of payment used by subscribers of different plans.

ENTITY RELATIONSHIP DIAGRAM

