Part 1

Database Schema & Tables

Using MongoDB means that our NoSQL database has a schema-less design. In the same manner, the data is not stored in tables but in collections. While our database itself does not enforce a specific schema, our data does follow a schema based on its structure. The following is the structure of our data:

1. Playlists.json

```
{
    "playlists": [
        {
            "name": "String",
            "lastModifiedDate": "ISO 8601",
            "items": [
                {
                    "track": {
                        "trackName": "String",
                        "artistName": "String",
                        "albumName": "String",
                        "trackUri": "URI"
                    },
                    "episode": null,
                    "localTrack": null
                }
            ],
            "description": null,
            "numberOfFollowers": "int"
        }
    ]
}
```

1. StreamingHistory.json

```
[
    {
        "endTime" : "ISO 8601",
        "artistName" : "String",
        "trackName" : "String",
        "msPlayed" : 179026
```

```
    }
]
```

1. SearchQueries.json

```
[
    {
        "platform" : "String",
        "searchTime" : "ISO 8601",
        "searchQuery" : "String",
        "searchInteractionURIs" : [
            "URI"
        ]
    }
]
```

Possible Data Distribution Plan

- Choose a Shard Key:
    - It should not be a field with monotonically increasing or decreasing values.
    - The field should have high cardinality.
- Determine Sharding Strategy:
    - Hash-Based Sharding: Distributes documents based on the hash value of the shard key. Useful for evenly distributing data but not for range-based queries.
    - Range-Based Sharding: Distributes documents based on ranges of shard key values. Good for queries that target certain ranges of data.
- Setup Shards:
    - Deploy multiple ‘mongod’ instances as shards.
    - Configure each shard with sufficient storage and processing power.
- Configure Config Servers
- Setup Multiple Query Routers (‘mongos’)

Data Insertion Mechanism

Our initial data insertion mechanism consists of loading the JSON files and using “insert_many” and “insert_one” based on the query to insert documents into their respective collections in our database. A “create” function was tailored to ensure efficient insertion of new data based on the structure of our existing data.
Further data insertion can be achieved in the same way, as well as other methods such as using “bulk_write”, using an upsert operation, and even using a streaming data platform for continuous data flow.

Data Retrieval

Query 1:

print("Sample Streaming History Document:", list(read(db.streaming_history)))

Output 1:

Sample Streaming History Document:

```
[{'_id': ObjectId('655a67930fe8a19acb69ddc8'), 'endTime': '2022-03-14 03:23', 'artistNam
e': 'Devon Again', 'trackName': 'Suburbia', 'msPlayed': 179026}]
```

Query 2:

print("First 5 docuents from Search Queries Collection:\n", list(read(db.search_queries, limit=5)))

Output 2:

First 5 docuents from Search Queries Collection:

```
[{'_id': ObjectId('655a65640fe8a19acb69b612'), 'platform': 'IPHONE', 'searchTime': '2022-0
1-05T21:28:53.795Z[UTC]', 'searchQuery': 'haruno', 'searchInteractionURIs': []}, {'_id': O
bjectId('655a65640fe8a19acb69b613'), 'platform': 'IPHONE', 'searchTime': '2022-01-10T01:5
6:52.118Z[UTC]', 'searchQuery': 'kick it', 'searchInteractionURIs': ['spotify:track:0yFq4G
HDqLfeceyAs63B3W']}, {'_id': ObjectId('655a65640fe8a19acb69b614'), 'platform': 'IPHONE',
'searchTime': '2022-01-10T01:56:59.754Z[UTC]', 'searchQuery': 'cherry bomb', 'searchIntera
ctionURIs': ['spotify:track:3o8QzWsiiqTUVgBZfHgF58']}, {'_id': ObjectId('655a65640fe8a19ac
b69b615'), 'platform': 'IPHONE', 'searchTime': '2022-01-10T01:57:28.812Z[UTC]', 'searchQue
ry': 'hot sauce', 'searchInteractionURIs': ['spotify:track:6B8MM3PVQtUbZLay7tP7er']}, {'_i
d': ObjectId('655a65640fe8a19acb69b616'), 'platform': 'IPHONE', 'searchTime': '2022-01-10T
02:06:39.721Z[UTC]', 'searchQuery': 'itzy', 'searchInteractionURIs': ['spotify:artist:2KC9
Qb60EaY0kW4eH68vr3']}]
```