

1) Business Scenario: What industry are you working in? Where might the data come from?

- The industry that we have chosen to work with is the Music Streaming Services Industry, specifically Spotify. Furthermore, the data that will be utilized is the user data collected by Spotify over a certain period of time as the data cannot be too large nor too small. Information about a person's listening habits, playlists, preferred genres, preferences, and maybe even their demographics are examples of user data. This information is gathered by Spotify in order to improve user experience, offer tailored suggestions, and advance its overall service.

2) Typical users: Who will need to interact with the data? Analysts? Customers? Managers?

- Algorithms: Machine learning algorithms examine user data to identify trends in user activity, offer tailored suggestions, and enhance the platform's overall user experience.
- Analysts: Spotify data analysts may work with this data to find patterns, gain insights, and provide suggestions based on data. They could be in charge of comprehending platform analytics as a whole, feature performance, and user engagement.
- Curators/Artists: Data is used by music curators and artists to comprehend consumer preferences, popular artists, and trending genres. By using this data, the platform's content curators may better curate playlists, suggest new releases, and maintain content relevancy for its user base.
- Users: Based on their listening habits and history in their Spotify data, users receive suggestions, customized playlists, and an optimized user experience. They can also interact with the data with third party services and Spotify Wrapped.

3) Representative questions (in plain language) your users will want to answer using your database: Recent Orders? Most popular product? Available rooms? Region with highest sales?

Our database can help answer questions that will be relevant to create internal and client-facing data reports. Spotify Wrapped is an example of a client-facing report that can be built from queries on this database. Wrapped describes user-specific scores, as well as scores comparing individual users to a broader audience. Query questions that are relevant here include:

- For each user, what is their most streamed song?
- Which genre did this user stream for the first time this year?
- What was this user's top ten most streamed genre?
- What is the percentage ranking of this user's streams for their most streamed artist, compared with other users who stream this artist?

Internal reports might help human or algorithmic playlist curators create suggestions. Queries like this may include:

- What is the most streamed song on a playlist?
- What is the average user playlist listening time?
- What are the top ten songs added to playlists in the last week?
- For each genre, which artist is trending for this week (received the most streams)?
- For each genre, which artist had the greatest increase in streams in the last week?

Other questions may include:

- For each record label, how many songs are represented on Spotify-curated playlists?
- What is the average number of playlists created by each user per year?

4) Anticipated volume of reads, inserts, and updates (imagining a real deployment): How many queries per day? How much new data is going in per day? How often does existing data change?

- The amount of queries being run per day will vary based on internal and client-facing needs. Internally, people such as analysts and curators will often run queries that are related to playlist trends, user interaction, and artist popularity multiple times a day whereas client-facing reports like Spotify wrapped, will generate queries much less frequently, like at the end of the or any other special occasions. The amount of new data going in per day is very abundant but also varies as it's dependent on user interaction and other external factors. This can range from users creating new playlists, adding songs to playlists, liking songs, new releases from artists, etc. In terms of how often data is changed, constant changes to existing data occur as it's also dependent on user interaction and other external factors. This ranges from users modifying their existing playlists, deleting songs, user preference changes, an update on artist streaming statistics, etc.

5) Consider which of your queries in 3), if any, must be very efficient: Where will performance matter?

- Building Spotify wrapped- fewer inserts post Dec 1, lots of queries across large amounts of fixed data, speed less relevant
- Most streamed song on the playlist: This query supports real-time recommendations and insights into trending music, necessitating swift execution.
- Average user playlist listening time: Important for understanding user engagement and for algorithmic recommendations, requiring efficient calculation.
- Top ten users with the most playlists and top ten artists represented on playlists: These queries are essential for personalizing user experience and promoting artists, thus need to be optimized for speed.

- Search queries for songs, artists, and playlists: As a core feature of the platform, search functionality must be highly responsive to ensure a good user experience.