Use Spotify Audio Features To Categorize Music

Practical Exam

Goal

Spotify provides an API for basic track information:

- **`ID`**, e.g. *2lEcSduKEXEK5KJ9hJzlCz*
- **`name`**, e.g. *Gloana Bauer (Teenage Dirtbag)*
- **`artist`**, e.g. *D‘ Hundskrippln*
- *…*

as well as related audio features:

- **`energy`**, e.g. *0.454*
  Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.
- **`speechiness`**, e.g. *0.0388*
  Speechiness detects the presence of spoken words in a track. The more exclu sively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
- **`loudness`**, e.g. *-8.758*
  The overall loudness of a track in decibels (dB). Loudness values are averaged.
- **`tempo`**, e.g. *97.532*
  The overall estimated tempo of a track in beats per minute (BPM)
- **`acousticness`**, *e.g. 0.268*
  A confidence measure from 0.0 to 1.0 of whether the track is acoustic.
- …

https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

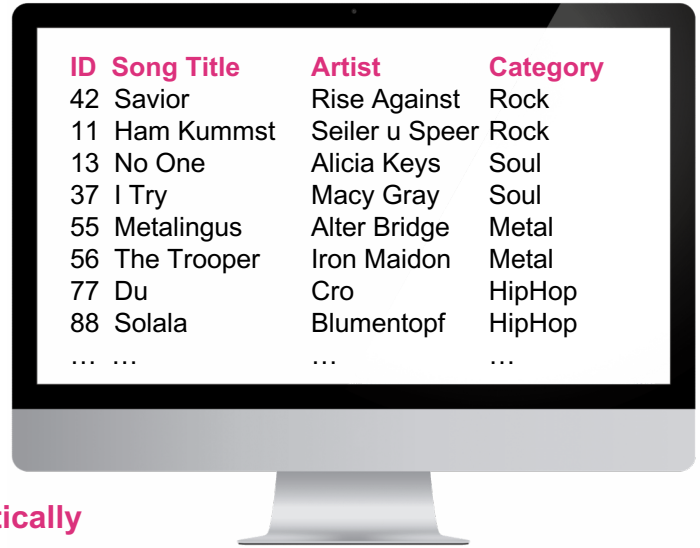
Goal

We want to make use of those audio features to automatically assign each track to a certain category:

Metal	***Classic***	***Rock***	***Vocal***
Electro	***Podcast***	***Soul***	***HipHop***

Workflow:

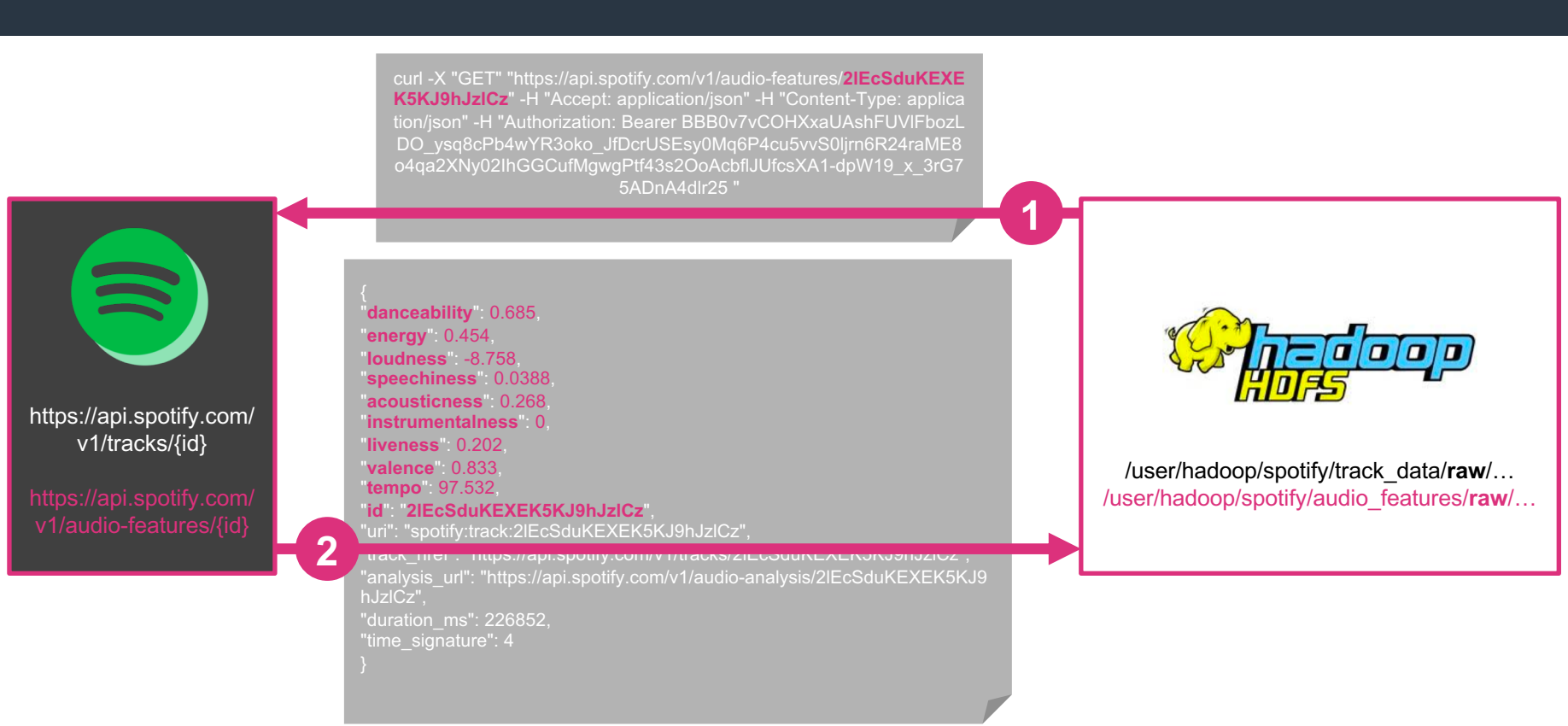
- **Query data** from Spotify API
- **Save raw data** (*JSON* files) to HDFS
- **Optimize**, **reduce** and **clean raw data** and save it to **final** directory on HDFS
- **Calculate categories** (*Metal, Classic, Rock, …*)
- **Join track information** and **audio festures** and **save** everything **to end-user database** (e.g. MySQL, MongoDB…)
- Provide a simple HTML Frontend which reads from end-user database and displays result
- The whole data workflow **must be implemented** within an ETL **workflow tool** (e.g. Pentaho Data Integration or Airflow) and **run automatically**



Dataflow: 1. Get Track Information

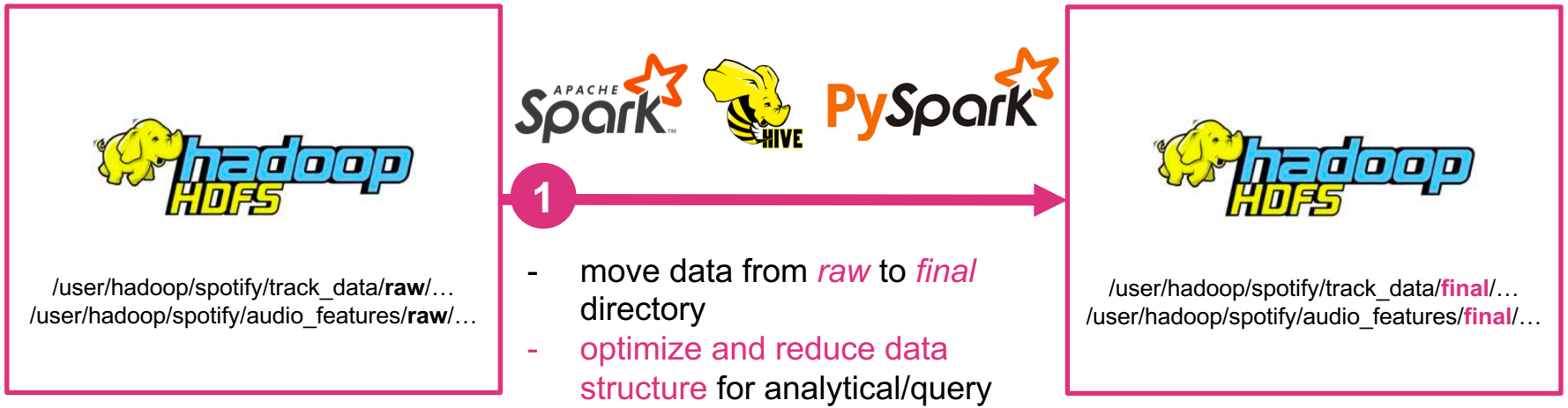
https://developer.spotify.com/console/get-track/

Dataflow: 2. Get Track Audio Features

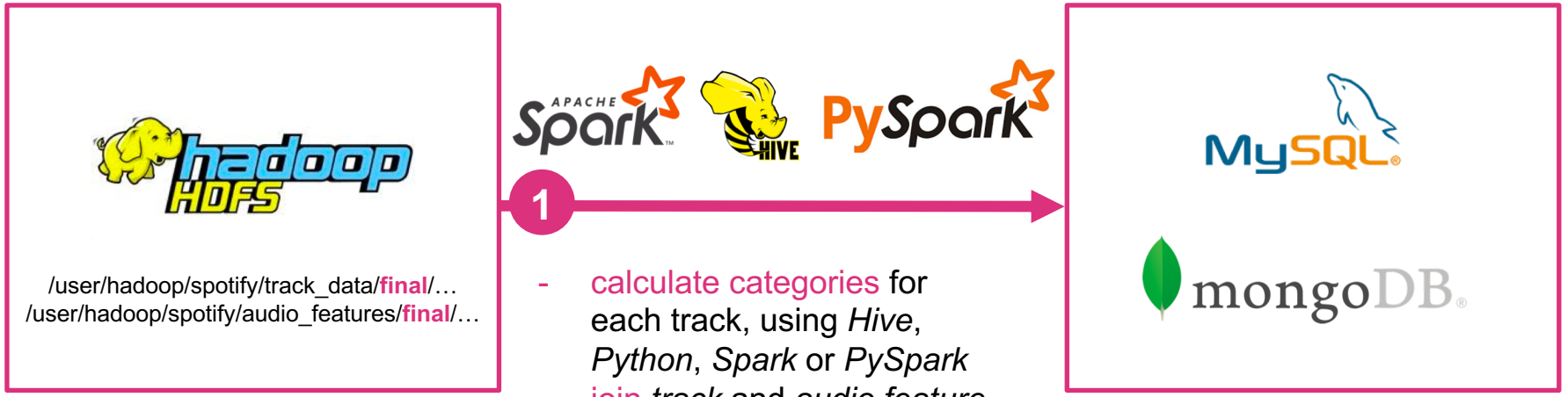


https://developer.spotify.com/console/get-audio-features-track/

Dataflow: 3. Raw To Final Transfer



Dataflow: 4. Run Analysis and Save Results

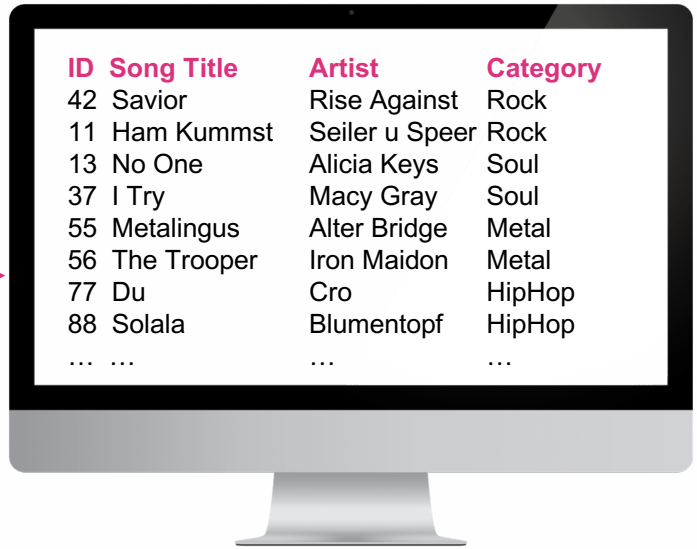


- calculate categories for each track, using *Hive*, *Python*, *Spark* or *PySpark*
- join *track* and *audio feature* data
- save everything to a end-user database (e.g. *MySQL*, *MongoDB*)

Dataflow: **5. Provide Simple Web Interface**

1

- Create a simple Website which reads and displays data from end-user database



ID	Song Title	Artist	Category
42	Savior	Rise Against	Rock
11	Ham Kummst	Seiler u Speer	Rock
13	No One	Alicia Keys	Soul
37	I Try	Macy Gray	Soul
55	Metalingus	Alter Bridge	Metal
56	The Trooper	Iron Maidon	Metal
77	Du	Cro	HipHop
88	Solala	Blumentopf	HipHop
…	…	…	…